Tight Security of Cascaded LRW2

Ashwin Jha and Mridul Nandi

Indian Statistical Institute, Kolkata, India
{ashwin.jha1991,mridul.nandi}@gmail.com

Abstract. At CRYPTO '12, Landecker et al. introduced the cascaded LRW2 (or CLRW2) construction, and proved that it is a secure tweakable block cipher up to roughly $2^{2n/3}$ queries. Recently, Mennink presented a distinguishing attack on CLRW2 in $2n^{1/2}2^{3n/4}$ queries. In the same paper, he discussed some non-trivial bottlenecks in proving tight security bound, i.e. security up to $2^{3n/4}$ queries. Subsequently, he proved security up to $2^{3n/4}$ queries for a variant of CLRW2 using 4-wise independent AXU assumption and the restriction that each tweak value occurs at most $2^{n/4}$ times. Moreover, his proof relies on a version of mirror theory which is yet to be publicly verified. In this paper, we resolve the bottlenecks in Mennink's approach and prove that the original CLRW2 is indeed a secure tweakable block cipher up to roughly $2^{3n/4}$ queries. To do so, we develop two new tools: First, we give a probabilistic result that provides improved bound on the joint probability of some special collision events; Second, we present a variant of Patarin's mirror theory in tweakable permutation settings with a self-contained and concrete proof. Both these results are of generic nature, and can be of independent interests. To demonstrate the applicability of these tools, we also prove tight security up to roughly $2^{3n/4}$ queries for a variant of DbHtS, called DbHtS-p, that uses two independent universal hash functions.

Keywords: LRW2, CLRW2, tweakable block cipher, mirror theory

1 Introduction

Tweakable Block Ciphers: A tweakable block cipher (or TBC for short) is a cryptographic primitive that has an additional public indexing parameter called tweak in addition to the usual secret key of a standard block cipher. This means that a tweakable block cipher, $\widetilde{E} : \mathcal{K} \times \mathcal{T} \times \mathcal{M} \to \mathcal{M}$, is a family of permutations on the plaintext/ciphertext space $\mathcal{M}$ indexed by two parameters: the secret key $k \in \mathcal{K}$ and the public tweak $t \in \mathcal{T}$. Liskov, Rivest, and Wagner formalized the concept of TBCs in their renowned work [1]. Tweakable block ciphers are more versatile than a standard block cipher and find a broad range of applications, most notably in authenticated encryption schemes, such as TAE [1], ΘCB [2] (TBC-based generalization of the OCB family [3,2,4]), PIV [5], COPA [6], SCT [7] (used in Deoxys [8,7]), AEZ [9] etc.; and message authentication codes, such as PMAC_TBC3K and PMAC_TBC1K [10], PMAC2x and PMACx [11], ZMAC [12], NaT and HaT [13], ZMAC+ [14], DoveMAC [15] etc. Apart from this TBCs have also been employed in encryption schemes [16,17,18,19,20,21].

Birthday-Bound Secure TBCs: Although there are some TBC constructions designed from scratch, notably Deoxys-BC [8] and Skinny [8,22], still the wide availability of secure and well-analyzed block ciphers make them perfect candidates for constructing TBCs. In [1], Liskov et al. proposed two constructions for TBCs based on a secure block cipher. The second construction, called LRW2, is defined as follows:

$$\mathsf{LRW2}((k,h),t,m) = E(k, m \oplus h(t)) \oplus h(t),$$

where E is a block cipher, k is the block cipher key, and h is an XOR universal hash function. The LRW2 construction is strongly related to the XEX construction by Rogaway [2], and its extensions by Chakraborty and Sarkar [23], Minematsu [24], and Granger et al. [25]. All these schemes are inherently birthday bound secure due to the internal hash XOR collisions, i.e. the adversary can choose approx. $2^{n/2}$ queries in such a way that there will be two queries (t, m) and (t', m') with $m \oplus h(t) = m' \oplus h(t')$. This leads to a simple distinguishing event $c \oplus c' = m \oplus m'$.

Beyond-the-Birthday Bound Secure TBCs: In [26], Landecker et al. first suggested the cascading of two independent LRW2 instances to get a beyond-the-birthday bound (BBB) secure TBC, called CLRW2, i.e.

$$\mathsf{CLRW2}((k_1,k_2,h_1,h_2),t,m) = \mathsf{LRW2}((k_2,h_2),t,\mathsf{LRW2}((k_1,h_1),t,m)).$$

They proved that CLRW2 is a secure TBC up to approx. $2^{2n/3}$ queries. Later on Procter [27] pointed out a flaw in the security proof of CLRW2. The proof was subsequently fixed by both Landecker et al. and Procter to recover the claimed security bounds. Lampe and Seurin [28] studied the $\ell \geq 2$ independent cascades of LRW2, and proved that it is secure up to approx. $2^{\frac{\ell n}{\ell+2}}$ queries. They further conjectured that the ℓ cascade is secure up to $2^{\frac{\ell n}{\ell+1}}$ queries. Recently, Mennink [29] showed a $2n^{1/2}2^{3n/4}$-query attack on CLRW2. In the same paper he also proved security up to $2^{3n/4}$ queries, albeit for a variant of CLRW2 with strong assumptions on the hash functions and restrictions on tweak repetitions.

All of the above constructions are proved to be secure in standard model. However, there are TBC constructions in public random permutation and ideal cipher model as well. In [13], Cogliati, Lampe and Seurin introduced the tweakable Even-Mansour construction and its cascaded variant. They showed that the two round construction is secure up to approx. $2^{2n/3}$ queries. A simple corollary of this result also gives security of CLRW2 up to $2^{2n/3}$ queries. The bound is tight in the ideal permutation model as one can simply fix the tweak and use the $2^{2n/3}$ queries attack on key alternating cipher by Bogdanov et al. [30]. Some notable BBB secure TBC constructions in the ideal cipher model include, Mennink's $\tilde{\mathsf{F}}[1]$ and $\tilde{\mathsf{F}}[2]$ [31,32], Wang et al. 32 constructions [33], and their generalization, called XHX, by Jha et al. [34]. All of these constructions are at most birthday bound secure in the sum of key size and block size.[1] Recently, Lee and Lee [35] proved that a two level cascade of XHX, called XHX2, achieves BBB security in terms of the sum of key size and block size.

[1] $\tilde{\mathsf{F}}[1]$, $\tilde{\mathsf{F}}[2]$, and Wang et al. constructions assume key size to be same as block size.

1.1 Recent Developments in the Analysis of CLRW2

In [29], Mennink presented an improved security analysis of CLRW2. The major contribution was an attack in approx. $n^{1/2}2^{3n/4}$ queries. The attack works by finding 4 queries (t, m_1, c_1), (t', m_2, c_2), (t, m_3, c_3), and (t', m_4, c_4) such that

$$\texttt{AltColl} \begin{cases} h_1(t) \oplus m_1 = h_1(t') \oplus m_2 & \wedge \quad h_2(t') \oplus c_2 = h_2(t) \oplus c_3 \\ h_1(t) \oplus m_3 = h_1(t') \oplus m_4 & \wedge \quad h_2(t') \oplus c_4 = h_2(t) \oplus c_1. \end{cases}$$

This leads to a simple distinguishing attack since, in case of CLRW2,

$$m_1 \oplus m_2 \oplus m_3 \oplus m_4 = 0 = c_1 \oplus c_2 \oplus c_3 \oplus c_4,$$

happens with probability 1, given AltColl holds. In contrast this happens with probability close to $1/2^n$ for an ideal tweakable random permutation.

Following on the insights from the attack, Mennink [29] also gives a security proof of the same order for a variant of CLRW2. Basically, the proof bounds the probability that the above given four equations hold. Additionally, inspired by [36], Patarin's mirror theory [37,38,39] is used which requires a bound on the probability of some more bad events. The major bottleneck in proving the security beyond $2^{2n/3}$ queries comes from two directions:

- First, there is no straightforward way of proving the upper bound of the probability of occurrence of AltColl to $\frac{q^4}{2^{3n}}$, where q is the number of queries. This is due to two reasons: (1) the adversary has full control over the tweak usages; and (2) the hash functions are just 2-wise independent XOR universal.
- Second, mirror theory was primarily developed to lower bound the number of solutions to equations arising for some random system which is trying to mimic a random function. This is not the case here, and as we will see in later sections, the mirror theory bound is directly dependent on tweak repetitions.

In order to bypass the two bottlenecks, following assumptions are made in [29]:

1. The hash functions are 4-wise independent AXU.
2. The maximum number of tweak repetitions is restricted to $2^{n/4}$.
3. A limited variant of mirror theory result is true for $q < 2^{3n/4}$.

Among the three assumptions, the first two are at least plausible. But the last assumption is questionable as barring certain restricted cases, the proof of mirror theory has many gaps which are still open or unproven, as has been noted in [40,41].

1.2 Contributions of this Work

In light of the above discussion, we revisit the proof strategy of [29] (see section 3), explicitly considering each of the issues. We show that all three assumptions used in [29] are dispensable. In order to do so, we develop some new tools which are described below:

1. *The Alternating Events Lemma:* We derive a generic tool (see section 4) to bound the probability of events of the form `AltColl`. In CLRW2 analysis only a special case is required, where the randomness comes from two independent universal hash functions.
2. *Mirror Theory in Tweakable Permutation Setting:* We adapt the mirror theory line of argument (see section 5) to get suitable bounds in tweakable permutation setting. This is a generalization of the existing mirror theory result in function setting.

Using the above mentioned tools we prove that CLRW2 is secure up to approx. $2^{3n/4}$ queries (see section 6). Our result, in combination with the attack in [29] (see supplementary material B), gives the tight (up to a logarithmic factor) security of CLRW2.

As a side-result on the application of our tools, we also prove tight security up to roughly $2^{3n/4}$ queries for a variant of DbHtS [42], called DbHtS-p, that uses two independent universal hash functions (see section 7).

Here, we explicitly remark that our bound on CLRW2 is not derivable from the recent result on XHX2 [35].

2 Preliminaries

Notational Setup: For $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, 2, \ldots, n\}$, $\{0,1\}^n$ denotes the set of bit strings of length n, and $\mathsf{Perm}(n)$ denotes the set of all permutations over $\{0,1\}^n$. For $n, \kappa \in \mathbb{N}$, $\mathsf{BPerm}(\kappa, n)$ denotes the set of all families of permutations $\pi_k := \pi(k, \cdot) \in \mathsf{Perm}(n)$, indexed by $k \in \{0,1\}^\kappa$. We sometimes extend this notation, whereby $\mathsf{BPerm}(\kappa, \tau, n)$ denotes the set of all families of permutations $\pi_{(k,t)}$, indexed by $(k,t) \in \{0,1\}^\kappa \times \{0,1\}^\tau$. For $n, r \in \mathbb{N}$, such that $n \geq r$, we define the falling factorial $(n)_r := n!/(n-r)! = n(n-1)\cdots(n-r+1)$.

For $q \in \mathbb{N}$, x^q denotes the q-tuple $(x_1, x_2, \ldots, x_q)$, and $\widehat{x^q}$ denotes the set $\{x_i : i \in [q]\}$. By an abuse of notation we also use x^q to denote the multiset $\{x_i : i \in [q]\}$ and $\mu(x^q, x')$ to denote the multiplicity of $x' \in x^q$. For a set $\mathcal{I} \subseteq [q]$ and a q-tuple x^q, $x^{\mathcal{I}}$ denotes the tuple $(x_i)_{i \in \mathcal{I}}$. For a pair of tuples x^q and y^q, (x^q, y^q) denotes the 2-ary q-tuple $((x_1, y_1), \ldots, (x_q, y_q))$. An n-ary q-tuple is defined analogously. For $q \in \mathbb{N}$, for any set $\mathcal{X}$, $(\mathcal{X})_q$ denotes the set of all q-tuples with distinct elements from $\mathcal{X}$. For $q \in \mathbb{N}$, a 2-ary tuple (x^q, y^q) is called permutation compatible, denoted $x^q \leftrightsquigarrow y^q$, if $x_i = x_j \iff y_i = y_j$. Extending notations, a 3-ary tuple (t^q, x^q, y^q) is called tweakable permutation compatible, denoted by $(t^q, x^q) \leftrightsquigarrow (t^q, y^q)$, if $(t_i, x_i) = (t_j, x_j) \iff (t_i, y_i) = (t_j, y_j)$. For any tuple $x^q \in \mathcal{X}^q$, and for any function $f : \mathcal{X} \to \mathcal{Y}$, $f(x^q)$ denotes the tuple $(f(x_1), \ldots, f(x_q))$. We use short hand notation $\exists^*$ to represent the phrase "there exists distinct".

We use the conventions: upper and lower case letters denote variables and values, respectively, and Serif font letters are used to denote random variables, unless stated otherwise. For a finite set $\mathcal{X}$, $\mathsf{X} \xleftarrow{\$} \mathcal{X}$ denotes the uniform and random sampling of X from $\mathcal{X}$.

2.1 Some Useful Inequalities

Definition 2.1. *For $r \geq s$, let $a = (a_i)_{i\in[r]}$ and $b = (b_j)_{j\in[s]}$ be two sequences over $\mathbb{N}$. We say that a compresses to b, if there exists a partition $\mathcal{P}$ of $[r]$ such that $\mathcal{P}$ contains exactly s cells, say $\mathcal{P}_1, \ldots, \mathcal{P}_s$, and $\forall i \in [s]$, $b_i = \sum_{j\in\mathcal{P}_i} a_j$.*

Proposition 1. *For $r \geq s$, let $a = (a_i)_{i\in[r]}$ and $b = (b_j)_{j\in[s]}$ be sequences over $\mathbb{N}$, such that a compresses to b. Then for any $n \in \mathbb{N}$, such that $2^n \geq \sum_{i=1}^{r} a_i$, we have $\prod_{i=1}^{r}(2^n)_{a_i} \geq \prod_{j=1}^{s}(2^n)_{b_j}$.*

In [34, Proof of Lemma 3], the authors refer to a variant of Proposition 1. We remark that, this variant [34, Fact 1] is in fact false. However, [34, Proof of Lemma 3] implicitly used Proposition 1, and hence stands correct.

Proposition 2. *For $r \geq 2$, let $c = (c_i)_{i\in[r]}$ and $d = (d_i)_{i\in[r]}$ be two sequences over $\mathbb{N}$. Let $a_1, a_2, b_1, b_2 \in \mathbb{N}$, such that $c_i \leq a_j$, $c_i + d_i \leq a_j + b_j$ for all $i \in [r]$ and $j \in [2]$, and $\sum_{i=1}^{r} d_i = b_1 + b_2$. Then, for any $n \in \mathbb{N}$, such that $a_j + b_j \leq 2^n$ for $j \in [2]$, we have $\prod_{i=1}^{r}(2^n - c_i)_{d_i} \geq (2^n - a_1)_{b_1}(2^n - a_2)_{b_2}$.*

Proposition 2 is quite intuitive, in the sense, that the starting value in each of the falling factorial term on the left is at least as much as the starting values on the right, and the total number of terms are same on both the sides. The formal proofs of Proposition 1 and 2 are given in supplementary material A.

2.2 (Tweakable) Block Ciphers and Random Permutations

A block cipher with key size κ and block size n is a family of permutations $E \in \mathsf{BPerm}(\kappa, n)$. For $k \in \{0,1\}^\kappa$, we denote $E_k(\cdot) := E(k, \cdot)$, and $E_k^{-1}(\cdot) := E^{-1}(k, \cdot)$. A tweakable block cipher with key size κ, tweak size τ and block size n is a family of permutations $\widetilde{E} \in \mathsf{BPerm}(\kappa, \tau, n)$. For $k \in \{0,1\}^\kappa$ and $t \in \{0,1\}^\tau$, we denote $\widetilde{E}_k(t, \cdot) := \widetilde{E}(k, t, \cdot)$, and $\widetilde{E}_k^{-1}(t, \cdot) := \widetilde{E}^{-1}(k, t, \cdot)$. Throughout this paper, we fix $\kappa, \tau, n \in \mathbb{N}$ as the key size, tweak size and block size, respectively, of the given (tweakable) block cipher.

We say that Π is an (ideal) random permutation on block space $\{0,1\}^n$ to indicate that $\Pi \leftarrow_\$ \mathsf{Perm}(n)$. Similarly, we say that $\widetilde{\Pi}$ is an (ideal) tweakable random permutation on tweak space $\{0,1\}^\tau$ and block space $\{0,1\}^n$ to indicate that $\widetilde{\Pi} \leftarrow_\$ \mathsf{BPerm}(\tau, n)$.

2.3 (T)SPRP 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. For instance, say an oracle gives bidirectional access (permutation P with inverse). If the adversary has made a forward call x and gets response $y = P(x)$. Then, making an inverse query y is redundant. Note that, such redundancies are necessary in certain security games, most notably in indifferentiability, where the adversary can use these redundancies to catch a

simulator. Let $\mathbb{A}(q,t)$ be the class of all non-trivial distinguishers limited to q oracle queries, and t computations.

(Tweakable) Strong Pseudorandom Permutation (SPRP): The SPRP advantage of distinguisher $\mathscr{A}$ against E instantiated with a key $\mathsf{K} \leftarrow\!\!\$\, \{0,1\}^\kappa$ is defined as

$$\mathbf{Adv}^{\mathsf{sprp}}_{E}(\mathscr{A}) = \mathbf{Adv}_{E^{\pm};\Pi^{\pm}}(\mathscr{A}) := \left|\Pr\left[\mathscr{A}^{E^{\pm}_{\mathsf{K}}} = 1\right] - \Pr\left[\mathscr{A}^{\Pi^{\pm}} = 1\right]\right|. \tag{1}$$

The SPRP security of E is defined as $\mathbf{Adv}^{\mathsf{sprp}}_{E}(q,t) := \max_{\mathscr{A}\in\mathbb{A}(q,t)} \mathbf{Adv}^{\mathsf{sprp}}_{E}(\mathscr{A})$.

Similarly, the TSPRP advantage of distinguisher $\mathscr{A}$ against $\widetilde{E}$ instantiated with a key $\mathsf{K} \leftarrow\!\!\$\, \{0,1\}^\kappa$ is defined as

$$\mathbf{Adv}^{\mathsf{tsprp}}_{\widetilde{E}}(\mathscr{A}) = \mathbf{Adv}_{\widetilde{E}^{\pm};\widetilde{\Pi}^{\pm}}(\mathscr{A}) := \left|\Pr\left[\mathscr{A}^{\widetilde{E}^{\pm}_{\mathsf{K}}} = 1\right] - \Pr\left[\mathscr{A}^{\widetilde{\Pi}^{\pm}} = 1\right]\right|. \tag{2}$$

The TSPRP security of $\widetilde{E}$ is defined as $\mathbf{Adv}^{\mathsf{tsprp}}_{\widetilde{E}}(q,t) := \max_{\mathscr{A}\in\mathbb{A}(q,t)} \mathbf{Adv}^{\mathsf{tsprp}}_{\widetilde{E}}(\mathscr{A})$.

2.4 The Expectation Method

Let $\mathscr{A}$ 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 $\mathscr{A}$'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_1 (res. Θ_0) the random transcript variable when $\mathscr{A}$ 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 $\mathscr{A}$ is deterministic, this probability depends only on the oracle $\mathcal{O}$ and the transcript ω. A transcript ω is said to be *attainable* if $\Pr[\Theta_0 = \omega] > 0$. The expectation method (stated below) is quite useful in obtaining improved bounds in many cases [43,44,45]. The H-coefficient technique due to Patarin [46] is a simple corollary of this result where the $\epsilon_{\mathsf{ratio}}$ is a constant function.

Lemma 2.1 (Expectation Method [43]). *Let Ω be the set of all transcripts. For some $\epsilon_{\mathsf{bad}} > 0$ and a non-negative function $\epsilon_{\mathsf{ratio}} : \Omega \to [0,\infty)$, suppose there is a set $\Omega_{\mathsf{bad}} \subseteq \Omega$ satisfying the following:*

- $\Pr[\Theta_0 \in \Omega_{\mathsf{bad}}] \le \epsilon_{\mathsf{bad}}$;
- *For any $\omega \notin \Omega_{\mathsf{bad}}$, ω is attainable and* $\dfrac{\Pr[\Theta_1 = \omega]}{\Pr[\Theta_0 = \omega]} \ge 1 - \epsilon_{\mathsf{ratio}}(\omega)$.

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

$$\mathbf{Adv}_{\mathcal{O}_1;\mathcal{O}_0}(\mathscr{A}) \le \epsilon_{\mathsf{bad}} + \mathsf{Ex}\left[\epsilon_{\mathsf{ratio}}(\Theta_0)\right].$$

2.5 Patarin's Mirror Theory

In [37] Patarin defines Mirror theory as a technique to estimate the number of solutions of linear systems of equalities and linear non equalities in finite groups. In its most general case, the mirror theory proof is tractable up to the order of $2^{2n/3}$ security bound, but it readily becomes complex and extremely difficult to verify, as one aims for the optimal bound [40,41]. We remark here that this in no way suggests that the result is incorrect, and in future, we might even get some independent verifications of the result.

We restrict ourselves to the binary field $\mathbb{F}_2^n$ with $\oplus$ as the group operation. We will use the Mennink and Neves interpretation [36] of mirror theory. For ease of understanding and notational coherency, we sometimes use different parametrization and naming conventions. Let $q \geq 1$ and let $\mathcal{L}$ be the system of linear equations

$$\{e_1 : Y_1 \oplus V_1 = \lambda_1, \quad e_2 : Y_2 \oplus V_2 = \lambda_2, \quad \ldots, \quad e_q : Y_q \oplus V_q = \lambda_q\}$$

where Y^q and V^q are unknowns, and $\lambda^q \in (\{0,1\}^n)^q$ are knowns. In addition there are (in)equality restrictions on Y^q and V^q, which uniquely determine $\widehat{Y}^q$ and $\widehat{V}^q$. We assume that $\widehat{Y}^q$ and $\widehat{V}^q$, are indexed in an arbitrary order by the index sets $[q_Y]$ and $[q_V]$, where $q_Y = |\widehat{Y}^q|$ and $q_V = |\widehat{V}^q|$. This assumption is without any loss of generality as this does not affect the system $\mathcal{L}$. Given such an ordering, we can view $\widehat{Y}^q$ and $\widehat{V}^q$ as ordered sets $\{Y'_1, \ldots, Y'_{q_Y}\}$ and $\{V'_1, \ldots, V'_{q_V}\}$, respectively. We define two surjective index mappings:

$$\varphi_Y : \begin{cases} [q] \to [q_Y] \\ i \mapsto j \text{ if and only if } Y_i = Y'_j. \end{cases} \qquad \varphi_V : \begin{cases} [q] \to [q_V] \\ i \mapsto k \text{ if and only if } V_i = V'_k. \end{cases}$$

It is easy to verify that $\mathcal{L}$ is uniquely determined by $(\varphi_Y, \varphi_V, \lambda^q)$, and vice-versa. Consider a labeled bipartite graph $\mathcal{G}(\mathcal{L}) = ([q_Y], [q_V], \mathcal{E})$ associated with $\mathcal{L}$, where $\mathcal{E} = \{(\varphi_Y(i), \varphi_V(i), \lambda_i) : i \in [q]\}$, λ_i being the label of edge. Clearly, each equation in $\mathcal{L}$ corresponds to a unique labeled edge (assuming no duplicate equations). We give three definitions with respect to the system $\mathcal{L}$ using $\mathcal{G}(\mathcal{L})$.

Definition 2.2 (cycle-freeness). *$\mathcal{L}$ is said to be cycle-free if and only if $\mathcal{G}(\mathcal{L})$ is acyclic.*

Definition 2.3 ($\xi_{\max}$-component). *Two distinct equations (or unknowns) in $\mathcal{L}$ are said to be in the same component if and only if the corresponding edges (res. vertices) in $\mathcal{G}(\mathcal{L})$ are in the same component. The size of any component $\mathcal{C}$ in $\mathcal{L}$, denoted $\xi(\mathcal{C})$, is the number of vertices in the corresponding component of $\mathcal{G}(\mathcal{L})$, and the maximum component size is denoted by $\xi_{\max}(\mathcal{L})$ (or simply $\xi_{\max}$).*

Definition 2.4 (non-degeneracy). *$\mathcal{L}$ is said to be non-degenerate if and only if there does not exist a path of even length at least 2 in $\mathcal{G}(\mathcal{L})$ such that the labels along the edges on this path sum up to zero.*

Theorem 2.1 (Fundamental Theorem of Mirror Theory [37]). *Let $\mathcal{L}$ be a system of equations over the unknowns $(\widehat{Y}^q, \widehat{V}^q)$, that is (i) cycle-free, (ii) non-degenerate, and (iii) $\xi_{\max}^2 \cdot \max\{q_Y, q_V\} \leq 2^n/67$. Then, the number of solutions $(y_1, \ldots, y_{q_Y}, v_1, \ldots, v_{q_V})$ of $\mathcal{L}$, denoted h_q, such that $y_i \neq y_j$ and $v_i \neq v_j$ for all $i \neq j$, satisfies*

$$h_q \geq \frac{(2^n)_{q_Y}(2^n)_{q_V}}{2^{nq}}. \tag{3}$$

A proof of this theorem is given in [37]. As mentioned before, the proof is quite involved with some claims remaining open or unproved. On the other hand, the same paper contains results for various other cases. For instance, for $\xi = 2$, several sub-optimal bounds have been shown. By sub-optimal, we mean that a factor of $(1-\epsilon)$, for some $\epsilon > 0$, is multiplied to the right hand side of Eq. (3). Inspired by this, we give the following terminology which will be useful in later references to mirror theory.

> For $\xi \geq 2$, $\epsilon > 0$, we write (ξ, ϵ)-restricted mirror theory theorem to denote the mirror theory result in which the number of solutions, h_q, of a system of equations with $\xi_{\max} = \xi$, satisfies $h_q \geq (1-\epsilon)\frac{(2^n)_{q_Y}(2^n)_{q_V}}{2^{nq}}$.

Mirror theory has been primarily used for bounding the pseudorandomness of sum of permutations [47,48,37,40] with respect to a random function. For instance, suppose we sample elements in $\widehat{Y}^q$ and $\widehat{V}^q$ as outputs of two independent random permutations Π_1 and Π_2, respectively, over q_Y and q_V distinct inputs, respectively. Let pr_1 be the probability of realizing the system of equations $\mathcal{L}$, and pr_0 be the probability of realizing the q-tuple λ^q through random function outputs over q distinct inputs. Then, it is easy to see that $\mathsf{pr}_1 = h_q/(2^n)_{q_Y}(2^n)_{q_V}$ and $\mathsf{pr}_0 = 1/2^{nq}$. Clearly, the above given lower bound on h_q implies that $\mathsf{pr}_1 \geq (1-\epsilon)\mathsf{pr}_0$. When combined with the H-coefficient technique, we get an ϵ term in the distinguishing advantage bound for sum of random permutations. Here ϵ can be viewed as the degree of deviation from random function behavior. This is precisely the reason that one finds terms of the form $(2^n)_{q_Y}(2^n)_{q_V}$ and 2^{nq} in mirror theory bounds. We refer the readers to [37,36] for a more detailed exposition on the aim and motivations behind mirror theory.

In [29], $(4, 3q/2^n)$-restricted mirror theory theorem is used. In section 5, we study the $(\xi, q^4/2^{3n})$ case, for $\xi \leq 2^n/2q$ and present a variant of mirror theory suitable for tweakable permutation scenario.

3 Revisiting Mennink's Improved Bound on CLRW2

We first describe the notion of ℓ-wise independent XOR universal hash functions as given in [29]. This notion will be used for the description of CLRW2 (for $\ell = 2$), as well as Mennink's improved bound on CLRW2 (for $\ell = 4$).

Definition 3.1. *For $\ell \geq 2$, $\epsilon \geq 0$, a family of functions $\mathcal{H} = \{h : \{0,1\}^\tau \to \{0,1\}^n\}$ is called an ℓ-wise independent XOR universal hash up to the bound*

ϵ, *denoted* ϵ-AXU_ℓ, *if for any* $j \in \{2, \ldots, \ell\}$, *any* $t^j \in (\{0,1\}^\tau)_j$ *and a* $\delta^{j-1} \in (\{0,1\}^n)^{j-1}$, *we have*

$$\Pr\left[\mathsf{H} \xleftarrow{\$} \mathcal{H} \ : \ \mathsf{H}(t_1) \oplus \mathsf{H}(t_2) = \delta_1, \ldots, \mathsf{H}(t_1) \oplus \mathsf{H}(t_j) = \delta_{j-1}\right] \leq \epsilon^{j-1}. \quad (4)$$

For $\ell = 2$, this is nothing but the notion of AXU hash functions, first introduced by Krawczyk [49] and later by Rogaway [50]. In [29], the author suggested a simple AXU_ℓ hash function family using finite field arithmetic for small domain ($\tau = n$). Basically, the hash function family is defined as follows

$$h(x) := \bigoplus_{i=1}^{\ell-1} h_i \odot x^i$$

for $h = (h_1, \ldots, h_{\ell-1})$, where $\odot$ denotes field multiplication operator with respect to some irreducible polynomial over the binary field $\mathbb{F}_2^n$. For $\ell = 2$, this yields the popular polyhash function. In general, this function requires $\ell - 1$ keys and $\ell - 1$ field multiplications to achieve 2^{-n}-AXU_ℓ. Alternatively, secure block ciphers can also be used to construct $(2^n - \ell + 1)^{-1}$-AXU_ℓ hash functions over sufficiently large domains.

3.1 Description of the Cascaded LRW2 Construction

Let $E \in \mathsf{BPerm}(\kappa, n)$ be a block cipher. Let $\mathcal{H}$ be a hash function family from $\{0,1\}^\tau$ to $\{0,1\}^n$. We define the tweakable block cipher $\mathsf{LRW2}[E, \mathcal{H}]$, based on the block cipher E and the hash function family $\mathcal{H}$, by the following mapping: $\forall (k, h, t, m) \in \{0,1\}^\kappa \times \mathcal{H} \times \{0,1\}^\tau \times \{0,1\}^n$,

$$\mathsf{LRW2}[E, \mathcal{H}](k, h, t, m) := E_k(m \oplus h(t)) \oplus h(t). \quad (5)$$

For $\ell \in \mathbb{N}$, the ℓ-round cascaded LRW2 construction, denoted $\mathsf{CLRW2}[E, \mathcal{H}, \ell]$, is a cascade of ℓ independent LRW2 instances, i.e. $\mathsf{CLRW2}[E, \mathcal{H}, \ell]$ is a tweakable block cipher, based on the block cipher E and the hash function family $\mathcal{H}$, defined as follows: $\forall (k^\ell, h^\ell, t, m) \in \{0,1\}^{\kappa\ell} \times \mathcal{H}^\ell \times \{0,1\}^\tau \times \{0,1\}^n$,

$$y_i := \begin{cases} \mathsf{LRW2}[E, \mathcal{H}](k_i, h_i, t, m) & \text{for } i = 1, \\ \mathsf{LRW2}[E, \mathcal{H}](k_i, h_i, t, y_{i-1}) & \text{otherwise.} \end{cases}$$

$$\mathsf{CLRW2}[E, \mathcal{H}, \ell](k^\ell, h^\ell, t, m) := y_\ell. \quad (6)$$

The 2-round CLRW2, was first analyzed by Landecker et al. [26], whereas the $\ell > 2$ case was studied by Lampe and Seurin [28]. Since we mainly focus on the $\ell = 2$ case, we use the nomenclatures, CLRW2 and cascaded LRW2, interchangeably with 2-round CLRW2. Figure 3.1 gives a pictorial description of the cascaded LRW2 construction. Throughout the rest of the paper, we use the notations from Figure 3.1 in context of CLRW2.

In [26] the CLRW2 construction was shown to be a BBB secure (upto $2^{2n/3}$ queries) TSPRP, provided the underlying block cipher is an SPRP, and the hash function families are AXU.

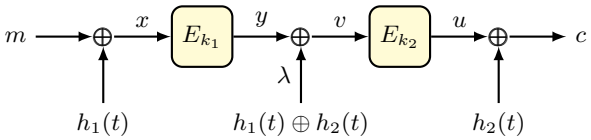

Fig. 3.1: The cascaded LRW2 construction.

3.2 Mennink's Proof Approach

The proof in [29] applies H-coefficient technique coupled with mirror theory. The main focus is to identify a suitable class of bad events on (x^q, u^q), where q is the number of queries, which makes mirror theory inapplicable. Crudely, the bad events correspond to cases where for some query there is no randomness left (in the sampling of y^q and v^q) in the ideal world. Given a good transcript, mirror theory is applied to bound the number of solutions of the system of equation $\{Y_i \oplus V_i = \lambda_i : i \in [q]\}$, where Y_i and V_i are unknowns satisfying $x^q \leftrightsquigarrow Y^q$ and $V^q \leftrightsquigarrow u^q$, and λ^q is fixed. The proof relies on three major assumptions:

Assumption 1. $\mathcal{H}$ is AXU_4 hash function family.

Assumption 2. For any $t' \in \{0,1\}^\tau$, $\mu_{t'} = \mu(t^q, t') \leq \gamma = 2^{n/4}$.

Assumption 3. $\left(4, \frac{3q}{2^n}\right)$-restricted mirror theory theorem is correct.

Transcript Graph: A graphical view on x^q and u^q was used to characterize all bad events. Basically, each transcript is mapped to a unique bipartite graph on x^q, u^q, as defined in Definition 3.2.

Definition 3.2 (Transcript Graph). *A transcript graph $\mathcal{G} = (\mathcal{X}, \mathcal{U}, \mathcal{E})$ associated with (x^q, u^q), denoted $\mathcal{G}(x^q, u^q)$, is defined as $\mathcal{X} := \{(x_i, 0) : i \in [q]\}$; $\mathcal{U} := \{(u_i, 1) : i \in [q]\}$; and $\mathcal{E} := \{((x_i, 0), (u_i, 1)) : i \in [q]\}$. We also associate the value $\lambda_i = h_1(t_i) \oplus h_2(t_i)$ with edge $((x_i, 0), (u_i, 1)) \in \mathcal{E}$.*

Note that the graph may not be simple, i.e. it can contain parallel edges. For all practical purposes we may drop the 0 and 1 for $(x, 0) \in \mathcal{X}$ and $(u, 1) \in \mathcal{U}$, as they can be easily distinguished from the context and notations. Further, for some $i, j \in [q]$, if $x_i = x_j$ (or $u_i = u_j$) , then they share the same vertex $x_i = x_j = x_{i,j}$ (or $u_i = u_j = u_{i,j}$). The event $x_i = x_j$ and $u_i = u_j$, although extremely unlikely, will lead to a parallel edge in $\mathcal{G}$. Finally each edge $(x_i, u_i) \in \mathcal{E}$ corresponds to a query index $i \in [q]$, so we can equivalently view (and call) the edge (x_i, u_i) as index i. Figure 3.2 gives an example graph for $\mathcal{G}$.

Bad Transcripts: A transcript graph $\mathcal{G}(x^q, u^q)$ is called bad if:

1. it has a cycle of size $= 2$.
2. it has two adjacent edges i and j such that $\lambda_i \oplus \lambda_j = 0$.
3. it has a component with number of edges ≥ 4.

All subgraphs in Figure 3.2, except the first two from left, are considered bad in [29]. Conditions 1 and 2 correspond to the cases which might lead to degeneracy

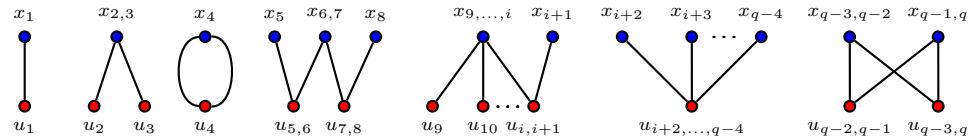

Fig. 3.2: A possible transcript graph $\mathcal{G}(x^q, u^q)$ associated with (x^q, u^q). Vertices in x^q are colored blue and vertices in u^q are colored red, for illustration only.

in the real world. Condition 3 may lead to a cycle of length ≥ 4 edges. The non-fulfillment of condition 1,2 and 3 satisfies the cycle-free and non-degeneracy properties required in mirror theory. It also bounds $\xi_{\max} \leq 4$. Condition 1 and 2 contribute small and insignificant terms and can be ignored from this discussion. We focus on the major bottleneck, i.e. condition 3. The subgraphs corresponding to condition 3 are given in Figure 3.3. Configuration (D), (E), and (F) are symmetric to (A), (B), and (C). So we can study (A), (B), and (C), and the other three can be similarly analyzed.

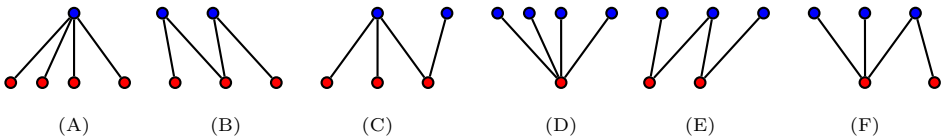

Fig. 3.3: Possible configuration of size = 4 edge subgraphs. Vertices in x^q are colored blue and vertices in u^q are colored red, and vertex labels are omitted for brevity.

Bottleneck 1: Bound on the probability of (A), (C), (D) and (F) — This can be divided into two parts:

(a) Configuration (A) arises for the event

$$\exists^* i, j, k, l \text{ such that } x_i = x_j = x_k = x_\ell.$$

This event is upper bounded to $q^4\epsilon^3$ using assumption 1 on hash functions. Similar argument holds for (D).

(b) Configuration (C) (similarly for F) arises for the event

$$\exists^* i, j, k, \ell \in [q] \text{ such that } x_i = x_j = x_k \wedge u_k = u_\ell.$$

In this case we can apply assumption 1 (even AXU_3 would suffice) to get an upper bound of $q^4\epsilon^3$.

Bottleneck 2: Bound on the probability of (B) — Configuration (B) arises for the event

$$\exists^* i, j, k, l \text{ such that } x_i = x_j \wedge u_j = u_k \wedge x_k = x_\ell.$$

This is probably the trickiest case, which requires assumption 2, i.e. restriction on tweak repetition. Specifically, consider the case $t_i = t_k$ and $t_j = t_\ell$. This is

precisely the case exploited in Mennink's attack on CLRW2 [29] (see supplementary material B). In this case for a fixed i, j, k, ℓ the probability is bounded by ϵ^2. There are at most q^2 choices for (i, j), at most $(\mu_{t_i} - 1)$ choices for k and a single choice for ℓ given i, j and k. Thus the probability is bounded by $q^2\gamma\epsilon^2$ (using assumption 2). Similar argument holds for (E).

Bottleneck 3: Mirror theory bound — The final hurdle is the use of mirror theory in computation of real world interpolation probability, which requires assumption 3. Yet another issue is the nature of the mirror theory bound. A straightforward application of mirror theory bound leads to a term of the form

$$\boxed{\frac{\prod_{t' \in \widehat{t}^q} (2^n)_{\mu_{t'}}}{2^{nq}}} (1 - O(q/2^n)) \,,$$

in the ratio of interpolation probabilities (as required for H-coefficient technique), where $\sum_{t' \in \widehat{t}^q} \mu_{t'} = q$. The boxed expression (particularly, the numerator in the expression) is of main interest. In the worst case, $\mu_{t'} = O(q)$, which gives a lower bound of the form $1 - q^2/2^n$ for the boxed expression. But using assumption 2, we get a lower bound of $1 - q\gamma/2^n$ as $\mu_{t'} \leq \gamma$.

Severity of the assumptions in [29]. Among the three assumptions, assumption 1 and 2 are plausible in the sense that real life use-cases exist for assumption 2 and practical instantiations are possible for assumption 1. Another point of note is the fact that $\gamma < 2^{n/4}$ is imposed due to bottleneck 3. Otherwise a better bound of $\gamma < 2^{n/2}$ could have been used. While assumption 1 and 2 are plausible to a large extent, assumption 3 is disputable. This is because no publicly verifiable proof exists for the generalized mirror theory. In fact, the proof for a special case of mirror theory also has some unproved gaps and mistakes. See Remark 1 for one such issue.

Although the proof in [29] requires the above mentioned assumptions, the proof approach seems quite simple and in some cases it highlights the bottlenecks in getting tight security. In the remainder of this paper, we aim to resolve all the bottlenecks discussed here, while relaxing all the assumptions made in [29]. Specifically, bottleneck 2 is resolved using the tools from section 4, and bottlenecks 1 and 3 are resolved using the tools from sections 4 and 5, and a careful application of the expectation method in section 6.

4 Results on (Multi)Collisions in Universal Hash

Let $\mathcal{H} = \{h \mid h : \mathcal{T} \to \mathcal{B}\}$ be a family of functions. A pair of distinct elements (t, t') from $\mathcal{T}$ is said to be colliding for a function $h \in \mathcal{H}$, if $h(t) = h(t')$. A family of functions $\mathcal{H} = \{h \mid h : \mathcal{T} \to \mathcal{B}\}$ is called an ϵ-universal hash if for all $t \neq t' \in \mathcal{T}$,

$$\Pr\left[\mathsf{H} \leftarrow\!\!\$\, \mathcal{H} \;:\; \mathsf{H}(t) = \mathsf{H}(t')\right] \leq \epsilon. \tag{7}$$

Throughout this section, we fix $t^q = (t_1, \ldots, t_q) \in (\mathcal{T})_q$. For a randomly chosen hash function $\mathsf{H} \xleftarrow{\$} \mathcal{H}$, the probability of having at least one colliding pair in t^q is at most $\binom{q}{2} \cdot \epsilon$. This is straightforward from the union bound.

4.1 The Alternating Collisions and Events Lemmata

Suppose $\mathcal{H}$ is an ϵ-universal hash and $\mathsf{H}_1, \mathsf{H}_2 \xleftarrow{\$} \mathcal{H}$ are two independently drawn universal hash functions. Then, by applying independence and union bound, we have

$$\Pr\left[\exists^* i, j, k \in [q],\ \mathsf{H}_1(t_i) = \mathsf{H}_1(t_j)\ \wedge\ \mathsf{H}_2(t_j) = \mathsf{H}_2(t_k)\right] \leq q(q-1)(q-2) \cdot \epsilon^2.$$

Now we go one step further. We would like to bound the probability of the following event:

$$\exists^* i, j, k, l \in [q],\ \mathsf{H}_1(t_i) = \mathsf{H}_1(t_j)\ \wedge\ \mathsf{H}_2(t_j) = \mathsf{H}_2(t_k)\ \wedge\ \mathsf{H}_1(t_k) = \mathsf{H}_1(t_l).$$

For any fixed distinct i, j, k and l, we cannot claim that the probability of the event $\mathsf{H}_1(t_i) = \mathsf{H}_1(t_j)\ \wedge\ \mathsf{H}_2(t_j) = \mathsf{H}_2(t_k)\ \wedge\ \mathsf{H}_1(t_k) = \mathsf{H}_1(t_l)$ is ϵ^3 as the first and last event are no longer independent. Now, we show how we can get an improved bound even in the dependent situation. In particular, we prove the following lemma.

Lemma 4.1 (Alternating Collisions Lemma). *Suppose $\mathsf{H}_1, \mathsf{H}_2 \xleftarrow{\$} \mathcal{H}$ are two independently drawn ϵ universal hash functions and $t^q \in (\mathcal{T})_q$. Then,*

$$\Pr\left[\exists^* i, j, k, l \in [q], \mathsf{H}_1(t_i) = \mathsf{H}_1(t_j) \wedge \mathsf{H}_1(t_k) = \mathsf{H}_1(t_l) \wedge \mathsf{H}_2(t_j) = \mathsf{H}_2(t_k)\right] \leq q^2 \epsilon^{1.5}.$$

Proof. For any $h \in \mathcal{H}$, we define the following useful set:

$$\mathcal{I}_h = \{(i, j) : h(t_i) = h(t_j)\}.$$

Let us denote the size of the above set by I_h. So, I_h is the number of colliding pairs for the hash functions h. We also define a set $\mathcal{H}_{\leq} = \{h : I_h \leq \frac{1}{\sqrt{\epsilon}}\}$ which collects all hash functions having a small number of colliding pairs. We denote the complement set by $\mathcal{H}_{>}$. Now, by using double counting of the set $\{(h, i, j) : h(t_i) = h(t_j)\}$ we get

$$\sum_h I_h \leq q(q-1) \cdot \epsilon \cdot |\mathcal{H}|. \tag{8}$$

Basically for every h, we have exactly I_h choices of (i, j) and so the size of the set $\{(h, i, j) : h(t_i) = h(t_j)\}$ is exactly $\sum_h I_h$. On the other hand, for any $1 \leq i < j \leq q$, there are at most $\epsilon \cdot |\mathcal{H}|$ hash functions h, such that (t_i, t_j) is a colliding pair for h. This follows from the definition of the universal hash function. From Eq. (8) and the definition of $\mathcal{H}_{\leq}$, we have

$$\frac{|\mathcal{H}_{>}|}{\sqrt{\epsilon}} + \sum_{h \in \mathcal{H}_{\leq}} I_h \ \leq \sum_h I_h \ \leq\ q(q-1) \cdot \epsilon \cdot |\mathcal{H}|. \tag{9}$$

Let $\mathtt{E}$ denote the event that there exists distinct i, j, k, l such that $\mathsf{H}_1(t_i) = \mathsf{H}_1(t_j) \wedge \mathsf{H}_1(t_k) = \mathsf{H}_1(t_l) \wedge \mathsf{H}_2(t_j) = \mathsf{H}_2(t_k)$. Now, we proceed to bound the probability of this event.

$$\begin{aligned}
\Pr[\mathtt{E}] &= \sum_h \Pr[\mathtt{E} \wedge \mathsf{H}_1 = h] \\
&= \sum_h \Pr[\mathsf{H}_1 = h] \times \Pr[\mathtt{E} \wedge \mathsf{H}_1 = h \mid \mathsf{H}_1 = h] \\
&\overset{1}{\leq} \sum_h \Pr[\mathsf{H}_1 = h] \times \min\{1, I_h^2 \cdot \epsilon\} \\
&= \Pr[\mathsf{H}_1 \in \mathcal{H}_>] + \sum_{h \in \mathcal{H}_\leq} \Pr[\mathsf{H}_1 = h] \cdot I_h^2 \cdot \epsilon \\
&\overset{2}{\leq} \frac{|\mathcal{H}_>|}{|\mathcal{H}|} + \sum_{h \in \mathcal{H}_\leq} \frac{I_h \cdot \sqrt{\epsilon}}{|\mathcal{H}|}. \\
&= \frac{\sqrt{\epsilon}}{|\mathcal{H}|} \times \left(\frac{|\mathcal{H}_>|}{\sqrt{\epsilon}} + \sum_{h \in \mathcal{H}_\leq} I_h \right) \\
&\overset{3}{\leq} q(q-1)\epsilon^{1.5}.
\end{aligned}$$

First, we justify inequality 1. Given $\mathsf{H}_1 = h$, the probability of the event $\mathtt{E}$ is same as the probability of the following event:

$$\exists^* (i, j), (k, l) \in \mathcal{I}_h,\ \mathsf{H}_2(t_j) = \mathsf{H}_2(t_k).$$

There are at most I_h^2 pairs of pairs and for each pair of pairs and the collision probability of $\mathsf{H}_2(t_j) = \mathsf{H}_2(t_k)$ is at most ϵ. So probability of the above event can be at most $\min\{1, I_h^2 \cdot \epsilon\}$. Now, we justify inequality 2 using two facts. First, $\mathsf{H}_1 \leftarrow\!\!\$\ \mathcal{H}$, i.e. $\Pr[\mathsf{H}_1 = h] = |\mathcal{H}|^{-1}$ for all $h \in \mathcal{H}$. Second, for all $h \in \mathcal{H}_\leq$, $I_h \leq 1/\sqrt{\epsilon}$. Inequality 3 follows from Eq. (9). □

Now, we generalize the above result for a more general setting. The proof of the result is similar to the previous proof and hence we skip it (given in supplementary material C).

Lemma 4.2 (Alternating Events Lemma). *Let $\mathsf{X}^q = (\mathsf{X}_1, \ldots, \mathsf{X}_q)$ be a q-tuple of random variables. Suppose for all $i < j \in [q]$, $\mathtt{E}_{i,j}$ are events associated with X_i and X_j, possibly dependent. Each event holds with probability at most ϵ. Moreover, for any distinct $i, j, k, l \in [q]$, $\mathtt{F}_{i,j,k,l}$ are events associated with X_i, X_j, X_k and X_l, which holds with probability at most ϵ'. Moreover, the collection of events $(\mathtt{F}_{i,j,k,l})_{i,j,k,l}$ is independent with the collection of event $(\mathtt{E}_{i,j})_{i,j}$. Then,*

$$\Pr[\exists^* i, j, k, l \in [q], \mathtt{E}_{i,j} \wedge \mathtt{E}_{k,l} \wedge \mathtt{F}_{i,j,k,l}] \leq q^2 \cdot \epsilon \cdot \sqrt{\epsilon'}$$

Note that, Lemma 4.1 is a direct corollary of the above Lemma (the event $\mathtt{E}_{i,j}$ denotes that (t_i, t_j) is a colliding pair of H_1 and $\mathtt{F}_{i,j,k,l}$ denotes that (t_j, t_k) is a colliding pair of H_2).

4.2 Expected Multicollisions in Universal Hash

Suppose $\mathcal{H}$ is an ϵ-universal hash, and $\mathsf{H} \xleftarrow{\$} \mathcal{H}$. Let $\mathsf{X}^q = \mathsf{H}(t^q)$. We define an equivalence relation $\sim$ on $[q]$ as: $\alpha \sim \beta$ if and only if $\mathsf{X}_\alpha = \mathsf{X}_\beta$ (i.e. $\sim$ is simply the multicollision relation). Let $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_r$ denote those equivalence classes of $[q]$ corresponding to $\sim$, such that $\nu_i = |\mathcal{P}_i| \geq 2$ for all $i \in [r]$. In the following lemma, we present a simple yet powerful result on multicollisions in universal hash functions.

Lemma 4.3. *Let* C *denote the number of colliding pairs in* X^q*. Then, we have*

$$\mathsf{Ex}\left[\sum_{i=1}^{r} {\nu_i}^2\right] \leq 2q^2\epsilon.$$

Proof. For $i \in [r]$, each of the $\binom{\nu_i}{2}$ pairs in $(\mathcal{P}_i)_2$ correspond to 1 colliding pair. And, each colliding pair belongs to $(\mathcal{P}_i)_2$ for some $i \in [r]$, as equality implies that the corresponding indices are related by $\sim$. Thus, we have

$$\sum_{i=1}^{r} {\nu_i}^2 = 2\mathsf{C} + \sum_{i=1}^{r} \nu_i \leq 4\mathsf{C}.$$

The result follows from the fact that $\mathsf{Ex}\left[\mathsf{C}\right] \leq \binom{q}{2}\epsilon$. □

Lemma 4.3 results in a simple corollary given below, which was independently proved in [51].

Corollary 4.1. *Let* $\nu_{\max} = \max\{\nu_i : i \in [r]\}$*. Then, for some* $a \geq 2$*, we have*

$$\Pr\left[\nu_{\max} \geq a\right] \leq \frac{2q^2\epsilon}{a^2}.$$

Proof. We have,

$$\Pr\left[\nu_{\max} \geq a\right] = \Pr\left[\nu_{\max}^2 \geq a^2\right] \leq \Pr\left[\mathsf{C} \geq a^2/4\right] \leq \frac{2q^2\epsilon}{a^2},$$

where the last inequality follows from Markov's inequality. □

Dutta et al. [52] proved a weaker[2] variant of Corollary 4.1, using an elegant combinatorial argumentation.

5 Mirror Theory in Tweakable Permutation Setting

As evident from bottleneck 3 of section 3.2, a straightforward application of mirror theory bound would lead to a sub-optimal bound. In order to circumvent

[2] The bound is $\frac{q^2\epsilon}{a}$.

this sub-optimality Mennink [29] used a restriction on tweak repetitions (assumption 2 of section 3.2). Specifically, a bound of the form $O(q/2^{3n/4})$ requires $\mu(t^q, t') < 2^{n/4}$ for all $t' \in \widehat{t^q}$, where t^q denotes the q-tuple of tweaks used in the q queries. In order to avoid this assumption, we need a different approach.

A closer inspection of the mirror theory proof reveals that we can actually avoid the restrictions on tweak repetitions. In fact, rather surprisingly, we will see that tweak repetitions are actually helpful in the sense that mirror theory bound is good. In the remainder of this section, we develop a modified version of mirror theory, apt for applications in tweakable permutation settings.

5.1 General Setup and Notations

Isolated and Star Components: In an edge-labeled bipartite graph $\mathcal{G} = (\mathcal{Y}, \mathcal{V}, \mathcal{E})$, an edge (y, v, λ) is called *isolated* edge if both y and v have degree 1. A component $\mathcal{S}$ of $\mathcal{G}$ is called star, if $\xi(\mathcal{S}) \geq 3$ and there exists a unique vertex v in $\mathcal{S}$ with degree $\xi(\mathcal{S}) - 1$. We call v the center of $\mathcal{S}$. Further, we call $\mathcal{S}$ a $\mathcal{Y}$-$\star$ (res. $\mathcal{V}$-$\star$) component if its center lies in $\mathcal{Y}$ (res. $\mathcal{V}$).

The System of Equation: Following the notations and definitions from section 2.5, consider a system of equation $\mathcal{L}$

$$\{e_1 : Y_1 \oplus V_1 = \lambda_1, \quad e_2 : Y_2 \oplus V_2 = \lambda_2, \quad \ldots, \quad e_q : Y_q \oplus V_q = \lambda_q\},$$

such that each component in $\mathcal{G}(\mathcal{L})$ is either an isolated edge or a star. Let c_1, c_2, and c_3 denote the number of components of isolated, $\mathcal{Y}$-$\star$, and $\mathcal{V}$-$\star$ types, respectively. Let q_1, q_2, and q_3 denote the number of equations of isolated, $\mathcal{Y}$-$\star$, and $\mathcal{V}$-$\star$ types, respectively. Therefore, $c_1 = q_1$.

Note that the equations in $\mathcal{L}$ can be arranged in any arbitrary order without affecting the number of solutions. For the sake of simplicity, we fix the ordering in such a way that all isolated edges occur first, followed by the star components.

Now, our goal is to give a lower bound on the number of solutions of $\mathcal{L}$, such that the Y_i' values are pairwise distinct and V_i' values are pairwise distinct. More formally, we aim to prove the following result.

Theorem 5.1. *Let $\mathcal{L}$ be the system of linear equations as described above with $q < 2^{n-2}$ and $\xi_{\max} q \leq 2^{n-1}$. Then, the number of tuples $(y_1, \ldots, y_{q_Y}, v_1, \ldots, v_{q_V})$ that satisfy $\mathcal{L}$, denoted h_q, such that $y_i \neq y_j$ and $v_i \neq v_j$, for all $i \neq j$, satisfies:*

$$h_q \geq \left(1 - \frac{13q^4}{2^{3n}} - \frac{2q^2}{2^{2n}} - \left(\sum_{i=1}^{c_2+c_3} \eta_{c_1+i}^2\right) \frac{4q^2}{2^{2n}}\right) \times \frac{(2^n)_{q_1+c_2+q_3} (2^n)_{q_1+q_2+c_3}}{\prod_{\lambda' \in \widehat{\lambda}^q} (2^n)_{\mu(\lambda^q, \lambda')}},$$

where $\eta_j = \xi_j - 1$ and ξ_j denotes the size (number of vertices) of the j-th component, for all $j \in [c_1 + c_2 + c_3]$.

We note here that the bound in Theorem 5.1 is parametrized in q and ξ. This is a bit different from the traditional mirror theory bounds. Further, we note that the bounds in Theorem 5.1, becomes $1 - O(q^4/2^{3n})$, when the value of $\sum_{i=1}^{c_2+c_3} \eta_{c_1+i}^2$

is $O(q^2/2^n)$. When we apply this result to CLRW2 and DbHtS-p, we can show that the expected value of the term is indeed $O(q^2/2^n)$ (a good time to revisit Lemma 4.3). Corollary 5.1, given below, is useful for random function setting.

Corollary 5.1. *Let $\mathcal{L}$ be the system of linear equations as described above with $q < 2^{n-2}$ and $\xi_{\max} q < 2^{n-1}$. Then, the number of tuples $(y_1, \ldots, y_{q_Y}, v_1, \ldots, v_{q_V})$ that satisfy $\mathcal{L}$, denoted h_q, such that $y_i \neq y_j$ and $v_i \neq v_j$, for all $i \neq j$, satisfies:*

$$h_q \geq \left(1 - \frac{13q^4}{2^{3n}} - \frac{2q^2}{2^{2n}} - \left(\sum_{i=1}^{c_2+c_3} \eta_{c_1+i}^2\right) \frac{4q^2}{2^{2n}}\right) \times \frac{(2^n)_{q_1+c_2+q_3}(2^n)_{q_1+q_2+c_3}}{2^{nq}},$$

where $\eta_j = \xi_j - 1$ and ξ_j denotes the size (number of vertices) of the j-th component, for all $j \in [c_1 + c_2 + c_3]$.

Note the difference between the expressions given in Theorem 5.1 and Corollary 5.1. Looking back at the discussion given towards the end of section 2.5, one can see the motivation behind the denominator given in Corollary 5.1. Since, we aim to apply mirror theory in tweakable permutation setting the denominator is changed accordingly in Theorem 5.1.

The proof of Theorem 5.1 uses an inductive approach similar to the one in [37]. We postpone the complete proof to supplementary material D.

6 Tight Security Bound of CLRW2

Based on the tools we developed in section 4 and 5, we now show that the CLRW2 construction achieves security up to the query complexity approximately $2^{3n/4}$. Given Mennink's attack [29] (see supplementary material B) in roughly these many queries we can conclude that the bound is tight.

Theorem 6.1. *Let $\kappa, \tau, n \in \mathbb{N}$ and $\epsilon > 0$. Let $E \in \mathsf{BPerm}(\kappa, n)$, and let $\mathcal{H}$ be an $(\{0,1\}^\tau, \{0,1\}^n, \epsilon)$-AXU hash function family. Consider*

$$\mathsf{CLRW2}[E, \mathcal{H}] : \{0,1\}^{2\kappa} \times \mathcal{H}^2 \times \{0,1\}^\tau \times \{0,1\}^n \to \{0,1\}^n.$$

For $q \leq 2^{n-2}$ and $t > 0$, the TSPRP *security of $\mathsf{CLRW2}[E, \mathcal{H}]$ against $\mathbb{A}(q,t)$ is given by*

$$\mathbf{Adv}^{\mathsf{tsprp}}_{\mathsf{CLRW2}[E,\mathcal{H}]}(q,t) \leq 2\mathbf{Adv}^{\mathsf{sprp}}_{E}(q,t') + \Delta,$$

where $t' = c(t + qt_{\mathcal{H}})$, $t_{\mathcal{H}}$ being the time complexity for computing the hash function $\mathcal{H}$, $c > 0$ is a constant depending upon the computation model, and

$$\Delta \leq 2q^2\epsilon^{1.5} + \frac{9q^4\epsilon^2}{2^n} + \frac{32q^4\epsilon}{2^{2n}} + \frac{13q^4}{2^{3n}} + 2q^2\epsilon^2 + \frac{2q^2}{2^{2n}}. \tag{10}$$

On putting $\epsilon = 1/2^n$, in Eq. (10) and further simplifying, we get

Corollary 6.1. *For $\epsilon = \frac{1}{2^n}$, we have*

$$\mathbf{Adv}^{\mathsf{tsprp}}_{\mathsf{CLRW2}[E,\mathcal{H}]}(q,t) \leq 2\mathbf{Adv}^{\mathsf{sprp}}_{E}(q,t') + \frac{54q^4}{2^{3n}} + \frac{2q^2}{2^{3n/2}} + \frac{4q^2}{2^{2n}}. \tag{11}$$

Specifically, the advantage bound is meaningful up to $q \approx 2^{\frac{3n}{4}-1.43}$ queries.

The proof of Theorem 6.1 employs the Expectation method coupled with an adaptation of $(2^n/2q, q^4/2^{3n})$-restricted mirror theory [37] in tweakable permutation settings. While our use of mirror theory is somewhat inspired by its recent use in [29], in contrast to [29], we apply a modified version of mirror theory and that too for a restricted subset of queries. The complete proof of Theorem 6.1 is given in the remainder of this section.

6.1 Initial Step

Consider the instantiation $\mathsf{CLRW2}[E_{\mathsf{K}_1}, E_{\mathsf{K}_2}, \mathsf{H}_1, \mathsf{H}_2]$ of $\mathsf{CLRW2}[E, \mathcal{H}]$, where K_1, K_2, H_1, H_2 are independent and $(\mathsf{K}_1, \mathsf{K}_2) \leftarrow_{\$} (\{0,1\}^{\kappa})^2$, $(\mathsf{H}_1, \mathsf{H}_2) \leftarrow_{\$} \mathcal{H}^2$. As the first step, we switch to the information-theoretic setting, i.e. we replace $(E_{\mathsf{K}_1}, E_{\mathsf{K}_2})$ with $(\Pi_1, \Pi_2) \leftarrow_{\$} \mathsf{Perm}(n)^2$. For the sake of simplicity, we write the modified instantiation $\mathsf{CLRW2}[\Pi_1, \Pi_2, \mathsf{H}_1, \mathsf{H}_2]$ as $\mathsf{CLRW2}$, i.e. without any parametrization. This switching is done via a standard hybrid argument that incurs a cost of $2\mathbf{Adv}^{\mathsf{sprp}}_{E}(q,t')$ where $t' = O(t + qt_{\mathcal{H}})$. Thus, we have

$$\mathbf{Adv}^{\mathsf{tsprp}}_{\mathsf{CLRW2}[E,\mathcal{H}]}(q,t) \leq 2\mathbf{Adv}^{\mathsf{sprp}}_{E}(q,t') + \mathbf{Adv}^{\mathsf{tsprp}}_{\mathsf{CLRW2}}(q). \tag{12}$$

So, in Eq. (12), we have to give an upper bound on $\mathbf{Adv}^{\mathsf{tsprp}}_{\mathsf{CLRW2}}(q)$. At this point, we are in the information-theoretic setting. In other words, we consider computationally unbounded distinguisher $\mathscr{A}$. Without loss of generality, we assume that $\mathscr{A}$ is deterministic and non-trivial. Under this setup, we are now ready to apply the expectation method.

6.2 Oracle Description

The two oracles of interest are: $\mathcal{O}_1$, the real oracle, that implements $\mathsf{CLRW2}$; and, $\mathcal{O}_0$, the ideal oracle, that implements $\widetilde{\Pi} \leftarrow_{\$} \mathsf{BPerm}(\tau, n)$. We consider an extended version of these oracles, the one in which they release some additional information. We use notations analogously as given in Figure 3.1 to describe the transcript generated by $\mathscr{A}$'s interaction with its oracle.

Description of the real oracle, $\mathcal{O}_1$: The real oracle $\mathcal{O}_1$ faithfully runs $\mathsf{CLRW2}$. We denote the transcript random variable generated by $\mathscr{A}$'s interaction with $\mathcal{O}_1$ by the usual notation Θ_1, which is a 10-ary q-tuple

$$(\mathsf{T}^q, \mathsf{M}^q, \mathsf{C}^q, \mathsf{X}^q, \mathsf{Y}^q, \mathsf{V}^q, \mathsf{U}^q, \lambda^q, \mathsf{H}_1, \mathsf{H}_2),$$

defined as follows: The initial transcript consists of $(\mathsf{T}^q, \mathsf{M}^q, \mathsf{C}^q)$, where for all $i \in [q]$:

– T_i: i-th tweak value, M_i: i-th plaintext value, C_i: i-th ciphertext value

where $\mathsf{C}^q = \mathsf{CLRW2}(\mathsf{T}^q, \mathsf{M}^q)$. At the end of the query-response phase $\mathcal{O}_1$ releases some additional information $(\mathsf{X}^q, \mathsf{Y}^q, \mathsf{V}^q, \mathsf{U}^q, \lambda^q, \mathsf{H}_1, \mathsf{H}_2)$, where for all $i \in [q]$:

– $(\mathsf{X}_i, \mathsf{Y}_i)$: i-th input-output pair for Π_1,
– $(\mathsf{V}_i, \mathsf{U}_i)$: i-th input-output pair for Π_2,
– λ_i: i-th internal masking, $\mathsf{H}_1, \mathsf{H}_2$: the hash keys.

Note that X^q, U^q, and λ^q are completely determined by the hash keys $\mathsf{H}_1, \mathsf{H}_2$, and the initial transcript $(\mathsf{T}^q, \mathsf{M}^q, \mathsf{C}^q)$. But, we include them anyhow to ease the analysis.

Description of the ideal oracle, $\mathcal{O}_0$: The ideal oracle $\mathcal{O}_0$ has access to $\widetilde{\Pi}$. Since $\mathcal{O}_1$ releases some additional information, $\mathcal{O}_0$ must generate these values as well. The ideal transcript random variable Θ_0 is also a 10-ary q-tuple

$$(\mathsf{T}^q, \mathsf{M}^q, \mathsf{C}^q, \mathsf{X}^q, \mathsf{Y}^q, \mathsf{V}^q, \mathsf{U}^q, \lambda^q, \mathsf{H}_1, \mathsf{H}_2),$$

defined below. Note that we use the same notation to represent the variables of transcripts in the both world. However, the probability distributions of the these would be determined from their definitions. The initial transcript consists of $(\mathsf{T}^q, \mathsf{M}^q, \mathsf{C}^q)$, where for all $i \in [q]$:

– T_i: i-th tweak value, M_i: i-th plaintext value, C_i: i-th ciphertext value,

where $\mathsf{C}^q = \widetilde{\Pi}(\mathsf{T}^q, \mathsf{M}^q)$. Once the query-response phase is over $\mathcal{O}_0$ first samples $(\mathsf{H}_1, \mathsf{H}_2) \xleftarrow{\$} \mathcal{H}^2$ and computes $\mathsf{X}^q, \mathsf{U}^q, \lambda^q$, where for all $i \in [q]$:

– $\mathsf{X}_i := \mathsf{H}_1(\mathsf{T}_i) \oplus \mathsf{M}_i$, $\mathsf{U}_i := \mathsf{H}_2(\mathsf{T}_i) \oplus \mathsf{C}_i$, $\lambda_i := \mathsf{H}_1(\mathsf{T}_i) \oplus \mathsf{H}_2(\mathsf{T}_i)$.

This means that X^q, U^q, and λ^q are defined honestly. Given the partial transcript $\Theta_0' := (\mathsf{T}^q, \mathsf{M}^q, \mathsf{C}^q, \mathsf{X}^q, \mathsf{U}^q, \lambda^q, \mathsf{H}_1, \mathsf{H}_2)$ we wish to characterize the hash key $\mathsf{H} := (\mathsf{H}_1, \mathsf{H}_2)$ as good or bad. We write $\mathcal{H}_{\mathsf{bad}}$ for the set of bad hash keys, and $\mathcal{H}_{\mathsf{good}} = \mathcal{H}^2 \setminus \mathcal{H}_{\mathsf{bad}}$. We say that a hash key $\mathsf{H} \in \mathcal{H}_{\mathsf{bad}}$ (or H is bad) if and only if one of the following predicates is true:

1. $\mathtt{H}_1$: $\exists^* i, j \in [q]$ such that $\mathsf{X}_i = \mathsf{X}_j \wedge \mathsf{U}_i = \mathsf{U}_j$.
2. $\mathtt{H}_2$: $\exists^* i, j \in [q]$ such that $\mathsf{X}_i = \mathsf{X}_j \wedge \lambda_i = \lambda_j$.
3. $\mathtt{H}_3$: $\exists^* i, j \in [q]$ such that $\mathsf{U}_i = \mathsf{U}_j \wedge \lambda_i = \lambda_j$.
4. $\mathtt{H}_4$: $\exists^* i, j, k, \ell \in [q]$ such that $\mathsf{X}_i = \mathsf{X}_j \wedge \mathsf{U}_j = \mathsf{U}_k \wedge \mathsf{X}_k = \mathsf{X}_\ell$.
5. $\mathtt{H}_5$: $\exists^* i, j, k, \ell \in [q]$ such that $\mathsf{U}_i = \mathsf{U}_j \wedge \mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_k = \mathsf{U}_\ell$.
6. $\mathtt{H}_6$: $\exists k \geq 2^n/2q, \exists^* i_1, i_2, \ldots, i_k \in [q]$ such that $\mathsf{X}_{i_1} = \cdots = \mathsf{X}_{i_k}$.
7. $\mathtt{H}_7$: $\exists k \geq 2^n/2q, \exists^* i_1, i_2, \ldots, i_k \in [q]$ such that $\mathsf{U}_{i_1} = \cdots = \mathsf{U}_{i_k}$.

Case 1. H is bad: If the hash key H is bad, then Y^q and V^q values are sampled degenerately as $\mathsf{Y}_i = \mathsf{V}_i = 0$ for all $i \in [q]$. It means that we sample without maintaining any specific conditions, which may lead to inconsistencies.

Case 2. H is good: To characterize the transcript corresponding to a good hash key it will be useful to study a graph, similar to the one in section 3, associated with $(\mathsf{X}^q, \mathsf{U}^q)$. Specifically, we consider the random transcript graph

$\mathcal{G}(\mathsf{X}^q, \mathsf{U}^q)$ arising due to $\mathsf{H} \in \mathcal{H}_{\mathsf{good}}$. Lemma 6.1 and Figure 6.1 characterizes the different types of possible components in $\mathcal{G}(\mathsf{X}^q, \mathsf{U}^q)$. Note that, type-2, type-3, type-4, and type-5 graphs are the same as configuration (A), (D), (C), and (F) of Figure 3.3, for ≥ 4 edges. These graphs are considered as bad in [29], whereas we allow such components.

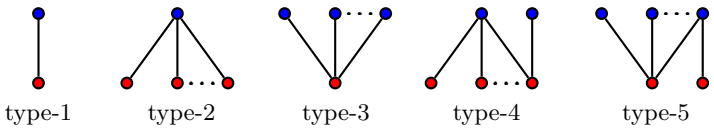

Fig. 6.1: Enumerating all possible types of components of a transcript graph corresponding to a good hash key: type-1 is the only possible component of size = 1 edge; type-2 and type-3 are $\mathcal{X}$-$\star$ and $\mathcal{U}$-$\star$ components, respectively; type-4 and type-5 are the only possible components that are not isolated or star (can have degree 2 vertices in both $\mathcal{X}$ and $\mathcal{U}$).

Lemma 6.1. *The transcript graph $\mathcal{G}$ corresponding to $(\mathsf{X}^q, \mathsf{U}^q)$ generated by a good hash key H has the following properties:*

1. *$\mathcal{G}$ is simple, acyclic and has no isolated vertices.*
2. *$\mathcal{G}$ has no two adjacent edges i and j such that $\lambda_i \oplus \lambda_j = 0$.*
3. *$\mathcal{G}$ has no component of size $> 2^n/2q$ edges.*
4. *$\mathcal{G}$ has no component such that it has 2 distinct degree 2 vertices in $\mathcal{X}$ or $\mathcal{U}$.*

In fact the all possible types of components of $\mathcal{G}$ are enumerated in Figure 6.1.

The proof of Lemma 6.1 is elementary and given in supplementary material E for the sake of completeness.

In what follows, we describe the sampling of Y^q and V^q when $\mathsf{H} \in \mathcal{H}_{\mathsf{good}}$. We collect the indices $i \in [q]$ corresponding to the edges in all type-1, type-2, type-3, type-4, and type-5 components, in the index sets $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$, $\mathcal{I}_4$, and $\mathcal{I}_5$, respectively. Clearly, the five sets are disjoint, and $[q] = \mathcal{I}_1 \sqcup \mathcal{I}_2 \sqcup \mathcal{I}_3 \sqcup \mathcal{I}_4 \sqcup \mathcal{I}_5$. Let $\mathcal{I} = \mathcal{I}_1 \sqcup \mathcal{I}_2 \sqcup \mathcal{I}_3$. Consider the system of equation

$$\mathcal{L} = \{Y_i \oplus V_i = \lambda_i \ : \ i \in \mathcal{I}\},$$

where $Y_i = Y_j$ (res. $V_i = V_j$) if and only if $\mathsf{X}_i = \mathsf{X}_j$ (res. $\mathsf{U}_i = \mathsf{U}_j$) for all $i, j \in [q]$. The solution set of $\mathcal{L}$ is precisely the set

$$\mathcal{S} = \{(y^{\mathcal{I}}, v^{\mathcal{I}}) : y^{\mathcal{I}} \leftrightsquigarrow \mathsf{X}^{\mathcal{I}} \wedge v^{\mathcal{I}} \leftrightsquigarrow \mathsf{U}^{\mathcal{I}} \wedge y^{\mathcal{I}} \oplus v^{\mathcal{I}} = \lambda^{\mathcal{I}}\}.$$

Given these definitions, the ideal oracle $\mathcal{O}_0$ samples $(\mathsf{Y}^q, \mathsf{V}^q)$ as follows:

- $(\mathsf{Y}^{\mathcal{I}}, \mathsf{V}^{\mathcal{I}}) \xleftarrow{\$} \mathcal{S}$, i.e. $\mathcal{O}_0$ uniformly samples one valid assignment from the set of all valid assignments.
- Let $\mathcal{G} \setminus \mathcal{I}$ denote the subgraph of $\mathcal{G}$ after the removal of edges and vertices corresponding to $i \in \mathcal{I}$. For each component $\mathcal{C}$ of $\mathcal{G} \setminus \mathcal{I}$:
  - Suppose $(\mathsf{X}_i, \mathsf{U}_i) \in \mathcal{C}$ corresponds to the edge in $\mathcal{C}$, where both X_i and U_i have degree ≥ 2. Then, $\mathsf{Y}_i \xleftarrow{\$} \{0,1\}^n$ and $\mathsf{V}_i = \mathsf{Y}_i \oplus \lambda_i$.

- For each edge $(\mathsf{X}_{i'}, \mathsf{U}_{i'}) \neq (\mathsf{X}_i, \mathsf{U}_i) \in \mathcal{C}$, either $\mathsf{X}_{i'} = \mathsf{X}_i$ or $\mathsf{U}_{i'} = \mathsf{U}_i$. Suppose, $\mathsf{X}_{i'} = \mathsf{X}_i$. Then, $\mathsf{Y}_{i'} = \mathsf{Y}_i$ and $\mathsf{V}_{i'} = \mathsf{Y}_{i'} \oplus \lambda_{i'}$. Now, suppose $\mathsf{U}_{i'} = \mathsf{U}_i$. Then, $\mathsf{V}_{i'} = \mathsf{V}_i$ and $\mathsf{Y}_{i'} = \mathsf{V}_{i'} \oplus \lambda_{i'}$.

At this point, $\Theta_0 = (\mathsf{T}^q, \mathsf{M}^q, \mathsf{C}^q, \mathsf{X}^q, \mathsf{Y}^q, \mathsf{V}^q, \mathsf{U}^q, \lambda^q, \mathsf{H}_1, \mathsf{H}_2)$ is completely defined. In this way we maintain both the consistency of equations of the form $\mathsf{Y}_i \oplus \mathsf{V}_i = \lambda_i$ (as in the case of real world), and the permutation consistency within each component, when $\mathsf{H} \in \mathcal{H}_{\mathsf{good}}$. However, there might be collisions among Y or V values from different components.

6.3 Definition and Analysis of Bad Transcripts

Given the description of the transcript random variable corresponding to the ideal oracle we can define the set of transcripts Ω as the set of all tuples $\omega = (t^q, m^q, c^q, x^q, y^q, v^q, u^q, \lambda^q, h_1, h_2)$, where $t^q \in (\{0,1\}^\tau)^q$; $m^q, c^q, y^q, v^q \in (\{0,1\}^n)^q$; $(h_1, h_2) \in \mathcal{H}^2$; $x^q = h_1(t^q) \oplus m^q$; $u^q = h_2(t^q) \oplus c^q$; $\lambda^q = h_1(t^q) \oplus h_2(t^q)$; and $(t^q, m^q) \leftrightsquigarrow (t^q, c^q)$.

Our bad transcript definition is inspired by two requirements:

1. Eliminate all x^q, u^q, and λ^q tuples such that both y^q and v^q are trivially restricted by way of linear dependence. For example, consider the condition $\mathtt{H}_2$. This leads to $y_i = y_j$, which would imply $v_i = y_i \oplus \lambda_i = y_j \oplus \lambda_j = v_j$. Assuming $i > j$, v_i is trivially restricted ($= v_j$) by way of linear dependence. This may lead to $u^q \not\leftrightsquigarrow v^q$ as u_i may not be equal to u_j.
2. Eliminate all x^q, u^q, y^q, v^q tuples such that $x^q \not\leftrightsquigarrow y^q$ or $u^q \not\leftrightsquigarrow v^q$.

Among the two, requirement 2 is trivial as $x^q \leftrightsquigarrow y^q$ and $u^q \leftrightsquigarrow v^q$ is always true for real world transcript. Requirement 1 is more of a technical one that helps in the ideal world sampling of y^q and v^q.

Bad Transcript Definition: We first define certain transcripts as bad depending upon the characterization of hash keys. Inspired by the ideal world description, we say that a hash key $(h_1, h_2) \in \mathcal{H}_{\mathsf{bad}}$ (or (h_1, h_2) is bad) if and only if the following predicate is true:

$$\mathtt{H}_1 \vee \mathtt{H}_2 \vee \mathtt{H}_3 \vee \mathtt{H}_4 \vee \mathtt{H}_5 \vee \mathtt{H}_6 \vee \mathtt{H}_7.$$

We say that ω is *hash induced bad* transcript, if $(h_1, h_2) \in \mathcal{H}_{\mathsf{bad}}$. We write this event as `BAD-HASH`, and by a slight abuse of notations,[3] we have

$$\mathtt{BAD\text{-}HASH} = \bigcup_{i=1}^{7} \mathtt{H}_i. \tag{13}$$

This takes care of the first requirement. For the second one we have to enumerate all the conditions which might lead to $x^q \not\leftrightsquigarrow y^q$ or $u^q \not\leftrightsquigarrow v^q$. Since we sample degenerately when the hash key is bad, the transcript is *trivially inconsistent* in this case. For good hash keys, if $x_i = x_j$ (or $u_i = u_j$) then we always have

[3] We use the notation $\mathtt{H}_i$ to denote the event that the predicate $\mathtt{H}_i$ is true.

$y_i = y_j$ (res. $v_i = v_j$); hence the inconsistency won't arise. So, given that the hash key is good, we say that ω is *sampling induced bad* transcript, if one of the following conditions is true:
for some $\alpha \in [5]$ and $\beta \in \{\alpha, \ldots, 5\}$, we have

- $\texttt{Ycoll}_{\alpha\beta}$: $\exists i \in \mathcal{I}_\alpha, j \in \mathcal{I}_\beta$, such that $x_i \neq x_j \wedge y_i = y_j$, and
- $\texttt{Vcoll}_{\alpha\beta}$: $\exists i \in \mathcal{I}_\alpha, j \in \mathcal{I}_\beta$, such that $u_i \neq u_j \wedge v_i = v_j$,

where $\mathcal{I}_i$ is defined as before in section 6.2. By varying α and β over all possible values, we get all 30 conditions which might lead to $x^q \nrightarrow y^q$ or $u^q \nrightarrow v^q$. Here we remark that some of these 30 conditions are never satisfied due to the sampling mechanism prescribed in section 6.2. These are $\texttt{Ycoll}_{11}$, $\texttt{Ycoll}_{12}$, $\texttt{Ycoll}_{13}$, $\texttt{Ycoll}_{22}$, $\texttt{Ycoll}_{23}$, $\texttt{Ycoll}_{33}$, $\texttt{Vcoll}_{11}$, $\texttt{Vcoll}_{12}$, $\texttt{Vcoll}_{13}$, $\texttt{Vcoll}_{22}$, $\texttt{Vcoll}_{23}$, and $\texttt{Vcoll}_{33}$. We listed them here only for the sake of completeness. We write the combined event that one of the 30 conditions hold as BAD-SAMP. Again by an abuse of notations, we have

$$\texttt{BAD-SAMP} = \bigcup_{\alpha\in[5],\beta\in\{\alpha,\ldots,5\}} (\texttt{Ycoll}_{\alpha\beta} \cup \texttt{Vcoll}_{\alpha\beta}) \,. \tag{14}$$

Finally, a transcript ω is called bad, i.e. $\omega \in \Omega_{\mathsf{bad}}$, if it is either a hash or a sampling induced bad transcript. All other transcripts are called good. It is easy to see that all good transcripts are attainable (as required in the H-coefficient technique or the expectation method).

Bad Transcript Analysis: We analyze the probability of realizing a bad transcript in the ideal world. By definition, this is possible if and only if one of BAD-HASH or BAD-SAMP occurs. So, we have

$$\begin{aligned}\epsilon_{\mathsf{bad}} = \Pr\left[\Theta_0 \in \Omega_{\mathsf{bad}}\right] &= \Pr_{\Theta_0}\left[\texttt{BAD-HASH} \cup \texttt{BAD-SAMP}\right] \\ &\leq \underbrace{\Pr_{\Theta_0}\left[\texttt{BAD-HASH}\right]}_{\epsilon_{\mathsf{hash}}} + \underbrace{\Pr_{\Theta_0}\left[\texttt{BAD-SAMP}\right]}_{\epsilon_{\mathsf{samp}}} . \end{aligned} \tag{15}$$

Lemma 6.2 upper bounds ϵ_{hash} to $2q^2\epsilon^2 + 2q^2\epsilon^{1.5} + 16q^4\epsilon 2^{-2n}$ and Lemma 6.3 upper bounds ϵ_{samp} to $9q^4\epsilon^2 2^{-n}$. Substituting these values in Eq. (15), we get

$$\epsilon_{\mathsf{bad}} \leq 2q^2\epsilon^2 + 2q^2\epsilon^{1.5} + \frac{16q^4\epsilon}{2^{2n}} + \frac{9q^4\epsilon^2}{2^n}. \tag{16}$$

Lemma 6.2. $\epsilon_{\mathsf{hash}} \leq 2q^2\epsilon^2 + 2q^2\epsilon^{1.5} + \dfrac{16q^4\epsilon}{2^{2n}}$.

Proof. Using Eq. (13) and (15), we have

$$\epsilon_{\mathsf{hash}} = \Pr\left[\mathtt{H}_1 \cup \mathtt{H}_2 \cup \mathtt{H}_3 \cup \mathtt{H}_4 \cup \mathtt{H}_5 \cup \mathtt{H}_6 \cup \mathtt{H}_7\right] \leq \sum_{i=1}^{7} \Pr\left[\mathtt{H}_i\right].$$

$\mathtt{H}_1$ is true if for some distinct i, j both $\mathsf{X}_i = \mathsf{X}_j$, and $\mathsf{U}_i = \mathsf{U}_j$. Now $\mathsf{T}_i = \mathsf{T}_j \implies \mathsf{M}_i \neq \mathsf{M}_j$. Hence $\mathsf{X}_i \neq \mathsf{X}_j$ and $\mathtt{H}_1$ is not true. So suppose $\mathsf{T}_i \neq \mathsf{T}_j$. Then for a

fixed i, j we get an upper bound of ϵ^2 as $\mathcal{H}$ is ϵ-AXU, and we have at most $\binom{q}{2}$ pairs of i, j. Thus, $\Pr[\mathtt{H}_1] \leq \binom{q}{2}\epsilon^2$. Following a similar line of argument one can bound $\Pr[\mathtt{H}_2] \leq \binom{q}{2}\epsilon^2$ and $\Pr[\mathtt{H}_3] \leq \binom{q}{2}\epsilon^2$.

In the remaining, we bound the probability of $\mathtt{H}_4$ and $\mathtt{H}_6$, while the probability of $\mathtt{H}_5$ and $\mathtt{H}_7$ can be bounded analogously. For any function $f : \{0,1\}^\tau \in \{0,1\}^n$, let $f' : \{0,1\}^\tau \times \{0,1\}^n \to \{0,1\}^n$ be defined as $f'(t, m) = f(t) \oplus m$. So $\mathsf{X}_i = \mathsf{H}'_1(\mathsf{T}_i, \mathsf{M}_i)$, and $\mathsf{U}_i = \mathsf{H}'_2(\mathsf{T}_i, \mathsf{C}_i)$, for all $i \in [q]$. It is easy to see that H'_b is ϵ-universal if H_b is ϵ-AXU for $b \in [2]$. Using the renewed description, $\mathtt{H}_4$ is true if for some distinct i, j, k, ℓ,

$$\mathsf{H}'_1(\mathsf{T}_i, \mathsf{M}_i) = \mathsf{H}'_1(\mathsf{T}_j, \mathsf{M}_j) \wedge \mathsf{H}'_2(\mathsf{T}_j, \mathsf{C}_j) = \mathsf{H}'_2(\mathsf{T}_k, \mathsf{C}_k) \wedge \mathsf{H}'_1(\mathsf{T}_k, \mathsf{M}_k) = \mathsf{H}'_1(\mathsf{T}_\ell, \mathsf{M}_\ell).$$

Since $(t_i, m_i) \neq (t_j, m_j)$ and $(t_i, c_i) \neq (t_j, c_j)$ for distinct i and j, we can apply the alternating collisions lemma of Lemma 4.1 to get $\Pr[\mathtt{H}_4] \leq q^2\epsilon^{1.5}$.
For $\mathtt{H}_6$, we have

$$\mathsf{X}_{i_1} = \mathsf{X}_{i_2} = \cdots = \mathsf{X}_{i_k},$$

where $k \geq 2^n/2q$. Since, $(t_{i_j}, m_{i_j}) \neq (t_{i_l}, m_{i_l})$ for all $j \neq l$, we can apply Corollary 4.1 with $a = 2^n/2q$ to get $\Pr[\mathtt{H}_6] \leq \frac{8q^4\epsilon}{2^{2n}}$. □

Lemma 6.3. $\epsilon_{\mathsf{samp}} \leq \dfrac{9q^4\epsilon^2}{2^n}$.

Proof. Using Eq. (14) and (15), we have

$$\begin{aligned}\epsilon_{\mathsf{samp}} &= \Pr\left[\bigcup_{\alpha\in[5],\beta\in\{\alpha,\ldots,5\}} (\mathtt{Ycoll}_{\alpha\beta} \cup \mathtt{Vcoll}_{\alpha\beta})\right] \\ &\leq \sum_{\alpha\in[5]} \sum_{\beta\in\{\alpha,\ldots,5\}} \left(\Pr[\mathtt{Ycoll}_{\alpha\beta}] + \Pr[\mathtt{Vcoll}_{\alpha\beta}]\right).\end{aligned}$$

We bound the probabilities of the events on the right hand side in groups as given below:

1. Bounding $\sum_{\alpha\in[3],\beta\in\{\alpha,\ldots,3\}} \Pr[\mathtt{Ycoll}_{\alpha\beta}] + \Pr[\mathtt{Vcoll}_{\alpha\beta}]$: Recall that the sampling of Y and V values is always done consistently for indices belonging to $\mathcal{I} = \mathcal{I}_1 \sqcup \mathcal{I}_2 \sqcup \mathcal{I}_3$. Hence,

$$\sum_{\alpha\in[3],\beta\in\{\alpha,\ldots,3\}} \Pr[\mathtt{Ycoll}_{\alpha\beta}] + \Pr[\mathtt{Vcoll}_{\alpha\beta}] = 0, \tag{17}$$

2. Bounding $\sum_{\alpha\in[3],\beta\in\{4,5\}} \Pr[\mathtt{Ycoll}_{\alpha\beta}] + \Pr[\mathtt{Vcoll}_{\alpha\beta}]$: Let's consider the event $\mathtt{Ycoll}_{14}$, which translates to there exist indices $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_4$ such that $\mathsf{X}_i \neq \mathsf{X}_j \wedge \mathsf{Y}_i = \mathsf{Y}_j$. Since $j \in \mathcal{I}_4$, there must exist $k, \ell \in \mathcal{I}_4 \setminus \{j\}$, such that one of the following happens

$$\mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_k = \mathsf{U}_\ell$$

$$\mathsf{U}_j = \mathsf{U}_k \wedge \mathsf{X}_k = \mathsf{X}_\ell$$
$$\mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_j = \mathsf{U}_\ell.$$

We analyze the first case, while the other two cases can be similarly bounded. To bound the probability of $\texttt{Ycoll}_{14}$, we can thus look at the joint event

$$\mathtt{E}: \ \exists i \in \mathcal{I}_1, \exists^* j,k,\ell \in \mathcal{I}_4, \text{ such that } \mathsf{Y}_i = \mathsf{Y}_j \wedge \mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_k = \mathsf{U}_\ell.$$

Note that the event $\mathsf{Y}_i = \mathsf{Y}_j$ is independent of $\mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_k = \mathsf{U}_\ell$, as both Y_i and Y_j are sampled independent of the hash key. Thus, we get

$$\begin{aligned}\Pr[\mathtt{E}] &= \Pr[\exists i \in \mathcal{I}_1, \exists^* j,k,\ell \in \mathcal{I}_4, \text{ such that } \mathsf{Y}_i = \mathsf{Y}_j \wedge \mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_k = \mathsf{U}_\ell] \\ &\le \sum_{i \in \mathcal{I}_1} \sum_{j<k<\ell \in \mathcal{I}_4} \Pr[\mathsf{Y}_i = \mathsf{Y}_j] \times \Pr[\mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_k = \mathsf{U}_\ell] \\ &\le q\binom{q}{3}\frac{\epsilon^2}{2^n},\end{aligned}$$

where the last inequality follows from the uniform randomness of Y_j and the AXU property of H_1 and H_2. The probability of the other two cases are similarly bounded to $q\binom{q}{3}\frac{\epsilon^2}{2^n}$, whence we get

$$\Pr[\texttt{Ycoll}_{14}] \le 3q\binom{q}{3}\frac{\epsilon^2}{2^n}.$$

We can bound the probabilities of $\texttt{Ycoll}_{24}$, $\texttt{Ycoll}_{34}$, $\texttt{Ycoll}_{\alpha 5}$, $\texttt{Vcoll}_{\alpha 4}$, and $\texttt{Vcoll}_{\alpha 5}$, for $\alpha \in [3]$, in a similar manner as in the case of $\texttt{Ycoll}_{14}$. So, we skip the argumentation for these cases, and summarize the probability for this group as

$$\sum_{\alpha \in [3], \beta \in \{4,5\}} \Pr[\texttt{Ycoll}_{\alpha\beta}] + \Pr[\texttt{Vcoll}_{\alpha\beta}] \le \frac{6q^4\epsilon^2}{2^n}. \tag{18}$$

3. Bounding $\sum_{\alpha \in \{4,5\}, \beta \in \{\alpha,5\}} \Pr[\texttt{Ycoll}_{\alpha\beta}] + \Pr[\texttt{Vcoll}_{\alpha\beta}]$: Consider the event $\texttt{Ycoll}_{44}$, which translates to there exists distinct indices $i,j \in \mathcal{I}_4$ such that $\mathsf{X}_i \neq \mathsf{X}_j \wedge \mathsf{Y}_i = \mathsf{Y}_j$. Here as $i,j \in \mathcal{I}_4$, there must exist $k,\ell \in \mathcal{I}_4 \setminus \{j\}$ such that one of the following happens

$$\mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_k = \mathsf{U}_\ell$$
$$\mathsf{U}_j = \mathsf{U}_k \wedge \mathsf{X}_k = \mathsf{X}_\ell$$
$$\mathsf{X}_j = \mathsf{X}_k \wedge \mathsf{U}_j = \mathsf{U}_\ell.$$

The analysis of these cases is similar to 2 above. So, we skip it and provide the final bound

$$\Pr[\texttt{Ycoll}_{44}] \le 3q\binom{q}{3}\frac{\epsilon^2}{2^n}.$$

The probabilities of all the remaining events in this group can be bounded in a similar fashion.

$$\sum_{\alpha\in\{4,5\},\beta\in\{\alpha,5\}} \Pr\left[\texttt{Ycoll}_{\alpha\beta}\right] + \Pr\left[\texttt{Vcoll}_{\alpha\beta}\right] \leq \frac{3q^4\epsilon^2}{2^n}. \tag{19}$$

The result follows by combining Eq. (17-19), followed by some algebraic simplifications. □

6.4 Good Transcript Analysis

From section 6.2, we know the types of components present in the transcript graph corresponding to a good transcript ω are exactly as in Figure 6.1. Let $\omega = (t^q, m^q, c^q, x^q, y^q, v^q, u^q, \lambda^q, h_1, h_2)$ be the good transcript at hand. From the bad transcript description of section 6.3, we know that for a good transcript $(t^q, m^q) \rightsquigarrow (t^q, c^q)$, $x^q \nrightarrow y^q$, $v^q \nrightarrow u^q$, and $y^q \oplus v^q = \lambda^q$.

We add some new parameters with respect to ω to aid our analysis of good transcripts. For $i \in [5]$, let $c_i(\omega)$ and $q_i(\omega)$ denote the number of components and number of indices (corresponding to the edges), respectively of type-i in ω. Note that $q_1(\omega) = c_1(\omega)$, $q_i(\omega) \geq 2c_i(\omega)$ for $i \in \{2,3\}$, and $q_i(\omega) \geq 3c_i(\omega)$ for $i \in \{4,5\}$. Obviously, for a good transcript $q = \sum_{i=1}^{5} q_i(\omega)$. For all these parameters, we will drop the ω parametrization whenever it is understood from the context.

Interpolation probability for the real oracle: In the real oracle, $(\mathsf{H}_1, \mathsf{H}_2) \xleftarrow{\$} \mathcal{H}^2$, Π_1 is called exactly $q_1 + c_2 + q_3 + 2c_4 + q_5 - c_5$ times and Π_2 is called exactly $q_1 + q_2 + c_3 + q_4 - c_4 + 2c_5$ times. Thus, we have

$$\Pr\left[\Theta_1 = \omega\right] = \frac{1}{|\mathcal{H}|^2} \times \frac{1}{(2^n)_{q_1+c_2+q_3+2c_4+q_5-c_5}} \times \frac{1}{(2^n)_{q_1+q_2+c_3+q_4-c_4+2c_5}}. \tag{20}$$

Interpolation probability for the ideal oracle: In the ideal oracle, the sampling is done in parts:

I. $\widetilde{\Pi}$ *sampling*: Let $(t'_1, t'_2, \cdots, t'_r)$ denote the tuple of distinct tweaks in t^q, and for all $i \in [r]$, let $a_i = \mu(t^q, t'_i)$, i.e. $r \leq q$ and $\sum_{i=1}^{r} a_i = q$. Then, we have

$$\Pr\left[\widetilde{\Pi}(t^q, m^q) = c^q\right] \leq \frac{1}{\prod_{i=1}^{r}(2^n)_{a_i}}.$$

II. *Hash key sampling*: The hash keys are sampled uniformly from $\mathcal{H}^2$, i.e. $\Pr\left[(\mathsf{H}_1, \mathsf{H}_2) = (h_1, h_2)\right] = \frac{1}{|\mathcal{H}|^2}$.

III. *Internal variables sampling*: The internal variables Y^q and V^q are sampled in two stages.

(A). *type*-1*, type*-2 *and type*-3 *sampling*: Recall the sets $\mathcal{I}_1$, $\mathcal{I}_2$, and $\mathcal{I}_3$, from section 6.3. Consider the system of equation

$$\mathcal{L} = \{\mathsf{Y}_i \oplus \mathsf{V}_i = \lambda_i \ : \ i \in \mathcal{I}\}.$$

Let $(\lambda'_1, \lambda'_2, \cdots, \lambda'_s)$ denote the tuple of distinct elements in $\lambda^{\mathcal{I}}$, and for all $i \in [s]$, let $b_i = \mu(\lambda^{\mathcal{I}}, \lambda'_i)$. From Figure 6.1 we know that $\mathcal{L}$ is cycle-free and non-degenerate. Further, $\xi_{\max}(\mathcal{L}) \leq 2^n/2q$, since the transcript is good. So, we can apply Theorem 5.1 to get a lower bound on the the number of valid solutions, $|\mathcal{S}|$ for $\mathcal{L}$. Using the fact that $(\mathsf{Y}^{\mathcal{I}}, \mathsf{V}^{\mathcal{I}}) \xleftarrow{\$} \mathcal{S}$, and Theorem 5.1, we have

$$\Pr\left[(\mathsf{Y}^{\mathcal{I}}, \mathsf{V}^{\mathcal{I}}) = (y^{\mathcal{I}}, v^{\mathcal{I}})\right] \leq \frac{\prod_{i=1}^{s}(2^n)_{b_i}}{\zeta(\omega)(2^n)_{q_1+c_2+q_3}(2^n)_{q_1+q_2+c_3}},$$

where

$$\zeta(\omega) = \left(1 - \frac{13q^4}{2^{3n}} - \frac{2q^2}{2^{2n}} - \left(\sum_{i=1}^{c_2+c_3} \eta_{c_1+i}^2\right)\frac{4q^2}{2^{2n}}\right),$$

(B). *type*-4*, and type*-5 *sampling*: For the remaining indices, one value is sampled uniformly for each of the components, i.e. we have

$$\Pr\left[\left(\mathsf{Y}^{[q]\setminus\mathcal{I}}, \mathsf{V}^{[q]\setminus\mathcal{I}}\right) = \left(y^{[q]\setminus\mathcal{I}}, v^{[q]\setminus\mathcal{I}}\right)\right] = \frac{1}{(2^n)^{c_4+c_5}}.$$

By combining I, II, III, and rearranging the terms, we have

$$\Pr\left[\Theta_0 = \omega\right] \leq \frac{1}{|\mathcal{H}|^2} \times \frac{1}{\zeta(\omega)} \times \frac{\prod_{i=1}^{s}(2^n)_{b_i}}{\prod_{i=1}^{r}(2^n)_{a_i}(2^n)_{p_1}(2^n)_{p_2}(2^n)^{c_4+c_5}}, \tag{21}$$

where $p_1 = q_1 + c_2 + q_3$, and $p_2 = q_1 + q_2 + c_3$.

6.5 Ratio of Interpolation Probabilities

On dividing Eq. (20) by Eq. (21), and simplifying the expression, we get

$$\begin{aligned}
\frac{\Pr\left[\Theta_1 = \omega\right]}{\Pr\left[\Theta_0 = \omega\right]} &\geq \zeta(\omega) \cdot \frac{\prod_{i=1}^{r}(2^n)_{a_i}}{\prod_{i=1}^{s}(2^n)_{b_i}(2^n - p_1 - c_4)_{c_4+q_5-c_5}(2^n - p_2 - c_5)_{q_4-c_4+c_5}} \\
&\overset{1}{\geq} \zeta(\omega) \cdot \frac{\prod_{i=1}^{r}(2^n)_{d_i}\prod_{i=1}^{r}(2^n - d_i)_{a_i-d_i}}{\prod_{i=1}^{s}(2^n)_{b_i}(2^n - p_1 - c_4)_{c_4+q_5-c_5}(2^n - p_2 - c_5)_{q_4-c_4+c_5}} \\
&\overset{2}{\geq} \zeta(\omega) \cdot \left.\frac{\prod_{i=1}^{r}(2^n - d_i)_{a_i-d_i}}{(2^n - p_1 - c_4)_{c_4+q_5-c_5}(2^n - p_2 - c_5)_{q_4-c_4+c_5}}\right\} A \\
&\overset{3}{\geq} \zeta(\omega).
\end{aligned} \tag{22}$$

At inequality 1, we rewrite the numerator such that $d_i = \mu(t^{\mathcal{I}}, t'_i)$ for $i \in [r]$. Further, $r \geq s$, as number of distinct internal masking values is at most the number of distinct tweaks, and $\widehat{t^{\mathcal{I}}}$ compresses to $\widehat{\lambda}^{\mathcal{I}}$. So using Proposition 1, we can justify inequality 2. At inequality 2, for $i \in \{2, 3, 4, 5\}$, $c_i(\omega) > 0$ if and only if $r \geq 2$. Also, $d_i \leq c_1 + c_2 + c_3 \leq p_1 + c_4$ and $d_i \leq p_2 + c_5$ for $i \in [r]$. Similarly, $a_i \leq c_1 + c_2 + c_3 + 2c_4 + 2c_5 \leq p_1 + 2c_4 + q_5 - c_5$, and $a_i \leq p_2 + q_4 - c_4 + 2c_5$. Also,

$\sum_{i=1}^{r} a_i - d_i = q_4 + q_5$. Thus, A satisfies the conditions given in Proposition 2, and hence $A \geq 1$. This justifies inequality 3.

We define $\epsilon_{\mathsf{ratio}} : \Omega \to [0, \infty)$ by the mapping

$$\epsilon_{\mathsf{ratio}}(\omega) = 1 - \zeta(\omega).$$

Clearly $\epsilon_{\mathsf{ratio}}$ is non-negative and the ratio of real to ideal interpolation probabilities is at least $1 - \epsilon_{\mathsf{ratio}}(\omega)$ (using Eq. (22)). Thus, we can use Lemma 2.1 to get

$$\mathbf{Adv}^{\mathsf{tsprp}}_{\mathsf{CLRW2}}(q) \leq \frac{2q^2}{2^{2n}} + \frac{13q^4}{2^{3n}} + \frac{4q^2}{2^{2n}} \mathsf{Ex}\left[\sum_{i=1}^{c_2+c_3} \eta^2_{c_1+i}\right] + \epsilon_{\mathsf{bad}}. \tag{23}$$

Let $\sim_1$ (res. $\sim_2$) be an equivalence relation over $[q]$, such that $\alpha \sim_1 \beta$ (res. $\alpha \sim_2 \beta$) if and only if $\mathsf{X}_\alpha = \mathsf{X}_\beta$ (res. $\mathsf{U}_\alpha = \mathsf{U}_\beta$). Now, each η_i random variable denotes the cardinality of some non-singleton equivalence class of $[q]$ with respect to either $\sim_1$ or $\sim_2$. Let $\mathcal{P}^1_1, \ldots, \mathcal{P}^1_r$ and $\mathcal{P}^2_1, \ldots, \mathcal{P}^2_s$ denote the non-singleton equivalence classes of $[q]$ with respect to $\sim_1$ and $\sim_2$, respectively. Further, for $i \in [r]$ and $j \in [s]$, let $\nu_i = |\mathcal{P}^1_i|$ and $\nu'_j = |\mathcal{P}^2_j|$. Then, we have

$$\begin{aligned}\mathsf{Ex}\left[\sum_{i=1}^{c_2+c_3} \eta^2_{c_1+i}\right] &\leq \mathsf{Ex}\left[\sum_{j=1}^{r} {\nu_j}^2\right] + \mathsf{Ex}\left[\sum_{k=1}^{s} {\nu'_k}^2\right] \\ &\leq 4q^2\epsilon. \end{aligned} \tag{24}$$

where the first inequality follows from the fact that H_1 and H_2 are independently sampled, and the second inequality follows from Lemma 4.3 and the fact that $\mathsf{H}_1, \mathsf{H}_2 \xleftarrow{\$} \mathcal{H}$. Theorem 6.1 follows from Eq. (12), (16), (23)-(24). □

7 Further Discussion

In this paper, our chief contribution is a tight (up to a logarithmic factor) security bound for the cascaded LRW2 tweakable block cipher. We developed two new tools: first, we provide a probabilistic result, called alternating collisions (events) lemma, that gives improved bounds for some special collision events, that are encountered frequently in BBB security analysis. Second, we adapt a restricted variant of mirror theory in tweakable permutations setting.

7.1 Applications of Alternating Events Lemma and Mirror Theory

The combination of alternating events lemma and mirror theory seem to have some nice applications. Here, we give some applications based on the Double-block Hash-then-Sum (or DbHtS) paradigm by Datta et al. [42]. The DbHtS paradigm is a variable input length pseudorandom function or PRF construction, based on a block cipher E and a hash function $\mathcal{H}$, which is defined as:
$\forall (k^2, h, m) \in \{0,1\}^{2\kappa} \times \mathcal{H} \times \{0,1\}^*$,

$$\mathsf{DbHtS}[E, \mathcal{H}](k^2, h, m) = \lambda = E_{k_1}(x) \oplus E_{k_2}(u),$$

where $\{0,1\}^*$ denotes the set of all bit strings, and $(x,u) = h(m)$.

PRF SECURITY: Let F be a keyed function family from $\{0,1\}^*$ to $\{0,1\}^n$ indexed by the key space $\{0,1\}^\kappa$. We define the PRF-advantage of an adversary $\mathscr{A}$ against F as,

$$\mathbf{Adv}^{\mathsf{prf}}_{F}(\mathscr{A}) = \left| \Pr_{\mathsf{K}} \left[\mathscr{A}^{F_{\mathsf{K}}} = 1 \right] - \Pr_{\Gamma} \left[\mathscr{A}^{\Gamma} = 1 \right] \right|,$$

where $\mathsf{K} \xleftarrow{\$} \{0,1\}^\kappa$, and Γ is a uniform random function chosen from the set of all functions from $\{0,1\}^*$ to $\{0,1\}^n$. The PRF security of F against any adversary class $\mathbb{A}(q,t)$ is defined analogously to SPRP and TSPRP security given in section 2.3.

Application 1: DbHtS-p— As a first application, we relax the DbHtS construction to DbHtS-p, where the hash function h is made up of independent universal hash functions h_1 and h_2, such that $h(m) = (h_1(m), h_2(m))$. This construction was also analyzed in [51], though they showed security up to $q \ll 2^{2n/3}$.

We show that DbHtS-p achieves higher security (i.e. security up to $q \ll 2^{3n/4}$). Further, the attack by Leurent et al. [53] in roughly $2^{3n/4}$ queries, seems to apply to DbHtS-p for algebraic hash functions. Thus, our bound is tight.

Theorem 7.1. *For $q \leq 2^{n-2}$ and $t > 0$, the* PRF *security of* DbHtS-p$[E,\mathcal{H}]$ *against $\mathbb{A}(q,t)$ is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{DbHtS\text{-}p}[E,\mathcal{H}]}(q,t) \leq 2\mathbf{Adv}^{\mathsf{prp}}_{E}(q,t') + \Delta,$$

where $t' = c(t + qt_{\mathcal{H}})$, $t_{\mathcal{H}}$ being the time complexity for computing the hash function $\mathcal{H}$, $c > 0$ is a constant depending upon the computation model, and

$$\Delta \leq 2q^2\epsilon^{1.5} + \frac{9q^4\epsilon^2}{2^n} + \frac{32q^4\epsilon}{2^{2n}} + \frac{13q^4}{2^{3n}} + q^2\epsilon^2 + \frac{q^2\epsilon}{2^n} + \frac{2q^2}{2^{2n}}. \tag{25}$$

Note that the PRP security game is similar to SPRP, except that the adversary is not given inverse access to the oracle. The proof of Theorem 7.1 is given in supplementary material F.

Application 2: DbHtS-f— The DbHtS-f is another relaxation of DbHtS, where the hash function h is made up of independent universal hash functions h_1 and h_2, and the finalization is done via keyed functions F_{k_1} and F_{k_2}, i.e., DbHtS-f$(m) = \lambda = F_{k_1}(x) \oplus F_{k_2}(u)$, where $x = h_1(m)$ and $u = h_2(m)$. We show that DbHtS-f is secure up to $q \ll 2^{3n/4}$.

Theorem 7.2. *For $q \leq 2^{n-2}$ and $t > 0$, the* PRF *security of* DbHtS-f$[F,\mathcal{H}]$ *against $\mathbb{A}(q,t)$ is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{DbHtS\text{-}f}[F,\mathcal{H}]}(q,t) \leq 2\mathbf{Adv}^{\mathsf{prf}}_{F}(q,t') + 2q^2\epsilon^{1.5} + q^2\epsilon^2, \tag{26}$$

where $t' = c(t + qt_{\mathcal{H}})$, $t_{\mathcal{H}}$ being the time complexity for computing the hash function $\mathcal{H}$, and $c > 0$ is a constant depending upon the computation model.

Proof. This can be argued using the previous line of research on sum of PRFs, starting from the work by Aiello and Venkatesan [54], followed by the works by Patarin et al.[55,56]. Basically, one can show that the sum of PRFs is a perfectly secure PRF if there is no "alternating cycles" in the inputs (see [54,55] for details). We can use the alternating collisions lemma to bound the probability of getting such alternating cycles. □

Modified Benes [54,55,56]: mBenes-f of [54] is a $2n$-bit to $2n$-bit PRF construction, which is defined as $\mathsf{mBenes\text{-}f}(a, b) := (e, f)$, where

$$c := F_{k_1}(a) \oplus b, \quad d := F_{k_2}(b) \oplus a, \quad e := F_{k_3}(c) \oplus F_{k_4}(d), \quad f := F_{k_5}(c) \oplus F_{k_6}(d),$$

where F_{k_i} are independently sampled PRFs. In [54], the authors conjectured that mBenes-f is secure up to $q \ll 2^n$. In [55,56], the authors have given a very high level sketch for proof of security up to $q \ll 2^{n-\epsilon}$ for all $\epsilon > 0$. Let us define the mappings $(a, b) \mapsto c$ and $(a, b) \mapsto d$ as functions h_1 and h_2. Then, it is easy to see that h_1 and h_2 are 2^{-n} universal hash functions. Hence, as a direct consequence of Theorem 7.2 above, one can argue that mBenes-f is secure up to $q \ll 2^{3n/4}$. Suppose mBenes-p denotes the natural variant of mBenes-f, when F_{k_i}'s are independently sampled PRPs. In the same vein as mBenes-f, mBenes-p can be shown to be secure up to $q \ll 2^{3n/4}$ as a direct consequence of Theorem 7.1 above.

7.2 Open Problems

We remark here that, the alternating events lemma is not applicable when the hash functions are dependent. Thus, we cannot apply it to other DbHtS instantiations, such as PMAC+ [57] and LightMAC+ [58], in a straightforward manner. It would be an interesting future work to somehow bypass the independence requirement of the alternating events lemma. Yet another future work could be to look for the repercussions of this result on the security of XHX2 [35] in both the ideal cipher and the standard model. Note that XHX2 in the standard model is same as 2-round cascade of XTX [59]. It seems that the bounds can be improved up to $\frac{3}{4}$-th of sum of block size and key size (or tweak size in the standard model). Our result does not seem to generalize to the cascaded LRW2 for $\ell > 2$, and it would be interesting to see some improved analysis on the generalized ℓ-round cascaded LRW2 for $\ell > 2$.

Update on the Security of PMAC+ and LightMAC+: In an independent and concurrent work [60] Kim et al. derived tight security bounds for several DbHtS-based MACs, which includes PMAC+ and LightMAC+. As a result, the exact security of these constructions is no longer an open problem. In order to derive tight security bounds, Kim et al. proposed an extension of mirror theory which is similar to Corollary 5.1 of this paper.

Acknowledgements

We thank the anonymous reviewers of EUROCRYPT 2019, CRYPTO 2019 and the Journal of Cryptology for their comments and suggestions. We also thank Bart Mennink for his comments and suggestions on an earlier version of this paper.

References

1. Liskov, M., Rivest, R.L., Wagner, D.A.: Tweakable block ciphers. J. Crypto. **24**(3) (2011) 588–613
2. Rogaway, P.: Efficient instantiations of tweakable blockciphers and refinements to modes OCB and PMAC. In: Advances in Cryptology - ASIACRYPT '04, Proceedings. (2004) 16–31
3. Rogaway, P., Bellare, M., Black, J., Krovetz, T.: OCB: a block-cipher mode of operation for efficient authenticated encryption. In: ACM Conference on Computer and Communications Security - ACM-CCS '01, Proceedings. (2001) 196–205
4. Krovetz, T., Rogaway, P.: The software performance of authenticated-encryption modes. In: Fast Software Encryption - FSE '11, Revised Selected Papers. (2011) 306–327
5. Shrimpton, T., Terashima, R.S.: A modular framework for building variable-input-length tweakable ciphers. In: Advances in Cryptology - ASIACRYPT '13, Proceedings, Part I. (2013) 405–423
6. Andreeva, E., Bogdanov, A., Luykx, A., Mennink, B., Tischhauser, E., Yasuda, K.: Parallelizable and authenticated online ciphers. In: Advances in Cryptology - ASIACRYPT '13, Proceedings, Part I. (2013) 424–443
7. Peyrin, T., Seurin, Y.: Counter-in-tweak: Authenticated encryption modes for tweakable block ciphers. In: Advances in Cryptology - CRYPTO '16, Proceedings, Part I. (2016) 33–63
8. Jean, J., Nikolic, I., Peyrin, T.: Tweaks and keys for block ciphers: The TWEAKEY framework. In: Advances in Cryptology - ASIACRYPT '14, Proceedings, Part II. (2014) 274–288
9. Hoang, V.T., Krovetz, T., Rogaway, P.: Robust authenticated-encryption AEZ and the problem that it solves. In: Advances in Cryptology - EUROCRYPT '15, Proceedings, Part I. (2015) 15–44
10. Naito, Y.: Full prf-secure message authentication code based on tweakable block cipher. In: Provable Security - ProvSec '15, Proceedings. (2015) 167–182
11. List, E., Nandi, M.: Revisiting full-prf-secure PMAC and using it for beyond-birthday authenticated encryption. In: Topics in Cryptology - CT-RSA '17, Proceedings. (2017) 258–274
12. Iwata, T., Minematsu, K., Peyrin, T., Seurin, Y.: ZMAC: A fast tweakable block cipher mode for highly secure message authentication. In: Advances in Cryptology - CRYPTO '17, Proceedings, Part III. (2017) 34–65
13. Cogliati, B., Lampe, R., Seurin, Y.: Tweaking even-mansour ciphers. In: Advances in Cryptology - CRYPTO '15, Proceedings, Part I. (2015) 189–208
14. List, E., Nandi, M.: ZMAC+ - an efficient variable-output-length variant of ZMAC. IACR Trans. Symmetric Cryptol. **2017**(4) (2017) 306–325
15. Grochow, T., List, E., Nandi, M.: Dovemac: A tbc-based PRF with smaller state, full security, and high rate. IACR Trans. Symmetric Cryptol. **2019**(3) (2019) 43–80

16. Minematsu, K.: Beyond-birthday-bound security based on tweakable block cipher. In: Fast Software Encryption - FSE '09, Revised Selected Papers. (2009) 308–326
17. Rogaway, P., Zhang, H.: Online ciphers from tweakable blockciphers. In: Topics in Cryptology - CT-RSA '11, Proceedings. (2011) 237–249
18. Forler, C., List, E., Lucks, S., Wenzel, J.: Poex: A beyond-birthday-bound-secure on-line cipher. Cryptography and Communications **10**(1) (2018) 177–193
19. Jha, A., Nandi, M.: On rate-1 and beyond-the-birthday bound secure online ciphers using tweakable block ciphers. Cryptography and Communications **10**(5) (2018) 731–753
20. Dutta, A., Nandi, M.: Tweakable HCTR: A BBB secure tweakable enciphering scheme. In: Progress in Cryptology - INDOCRYPT '18, Proceedings. (2018) 47–69
21. Bhaumik, R., List, E., Nandi, M.: ZCZ - achieving n-bit SPRP security with a minimal number of tweakable-block-cipher calls. In: Advances in Cryptology - ASIACRYPT '18, Proceedings, Part I. (2018) 336–366
22. Beierle, C., Jean, J., Kölbl, S., Leander, G., Moradi, A., Peyrin, T., Sasaki, Y., Sasdrich, P., Sim, S.M.: The SKINNY family of block ciphers and its low-latency variant MANTIS. In: Advances in Cryptology - CRYPTO '16, Proceedings, Part II. (2016) 123–153
23. Chakraborty, D., Sarkar, P.: A general construction of tweakable block ciphers and different modes of operations. IEEE Trans. Information Theory **54**(5) (2008) 1991–2006
24. Minematsu, K.: Improved security analysis of XEX and LRW modes. In: Selected Areas in Cryptography - SAC '06, Revised Selected Papers. (2006) 96–113
25. Granger, R., Jovanovic, P., Mennink, B., Neves, S.: Improved masking for tweakable blockciphers with applications to authenticated encryption. In: Advances in Cryptology - EUROCRYPT '16, Proceedings, Part I. (2016) 263–293
26. Landecker, W., Shrimpton, T., Terashima, R.S.: Tweakable blockciphers with beyond birthday-bound security. In: Advances in Cryptology - CRYPTO '12, Proceedings. (2012) 14–30
27. Procter, G.: A note on the CLRW2 tweakable block cipher construction. IACR Cryptology ePrint Archive **2014** (2014) 111
28. Lampe, R., Seurin, Y.: Tweakable blockciphers with asymptotically optimal security. In: Fast Software Encryption - FSE '13, Revised Selected Papers. (2013) 133–151
29. Mennink, B.: Towards tight security of cascaded LRW2. In: Theory of Cryptography - TCC '18, Proceedings, Part II. (2018) 192–222
30. Bogdanov, A., Knudsen, L.R., Leander, G., Standaert, F., Steinberger, J.P., Tischhauser, E.: Key-alternating ciphers in a provable setting: Encryption using a small number of public permutations - (extended abstract). In: Advances in Cryptology - EUROCRYPT '12, Proceedings. (2012) 45–62
31. Mennink, B.: Optimally secure tweakable blockciphers. In: Fast Software Encryption - FSE '15, Revised Selected Papers. (2015) 428–448
32. Mennink, B.: Optimally secure tweakable blockciphers. IACR Cryptology ePrint Archive **2015** (2015) 363
33. Wang, L., Guo, J., Zhang, G., Zhao, J., Gu, D.: How to build fully secure tweakable blockciphers from classical blockciphers. In: Advances in Cryptology - ASIACRYPT '16, Proceedings, Part I. (2016) 455–483
34. Jha, A., List, E., Minematsu, K., Mishra, S., Nandi, M.: XHX - A framework for optimally secure tweakable block ciphers from classical block ciphers and universal hashing. In: Progress in Cryptology - LATINCRYPT '17, Revised Selected Papers. (2017) 207–227

35. Lee, B., Lee, J.: Tweakable block ciphers secure beyond the birthday bound in the ideal cipher model. In: Advances in Cryptology - ASIACRYPT '18, Proceedings, Part I. (2018) 305–335
36. Mennink, B., Neves, S.: Encrypted davies-meyer and its dual: Towards optimal security using mirror theory. In: Advances in Cryptology - CRYPTO '17, Proceedings, Part III. (2017) 556–583
37. Patarin, J.: Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. IACR Cryptology ePrint Archive **2010** (2010) 287
38. Patarin, J.: Mirror theory and cryptography. Appl. Algebra Eng. Commun. Comput. **28**(4) (2017) 321–338
39. Nachef, V., Patarin, J., Volte, E.: Feistel Ciphers - Security Proofs and Cryptanalysis. Springer (2017)
40. Dai, W., Hoang, V.T., Tessaro, S.: Information-theoretic indistinguishability via the chi-squared method. In: Advances in Cryptology - CRYPTO '17, Proceedings, Part III. (2017) 497–523
41. Datta, N., Dutta, A., Nandi, M., Yasuda, K.: Encrypt or decrypt? to make a single-key beyond birthday secure nonce-based MAC. In: Advances in Cryptology - CRYPTO '18, Proceedings, Part I. (2018) 631–661
42. Datta, N., Dutta, A., Nandi, M., Paul, G.: Double-block hash-then-sum: A paradigm for constructing bbb secure prf. IACR Trans. Symmetric Cryptol. **2018**(3) (2018) 36–92
43. Hoang, V.T., Tessaro, S.: Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In: Advances in Cryptology - CRYPTO '16, Proceedings, Part I. (2016) 3–32
44. Hoang, V.T., Tessaro, S.: The multi-user security of double encryption. In: Advances in Cryptology - EUROCRYPT '17, Proceedings, Part II. (2017) 381–411
45. Guo, C., Wang, L.: Revisiting key-alternating feistel ciphers for shorter keys and multi-user security. In: Advances in Cryptology - ASIACRYPT '18, Proceedings, Part I. (2018) 213–243
46. Patarin, J.: Etude des Générateurs de Permutations Pseudo-aléatoires Basés sur le Schéma du DES. PhD thesis, Université de Paris (1991)
47. Hall, C., Wagner, D.A., Kelsey, J., Schneier, B.: Building prfs from prps. In: Advances in Cryptology - CRYPTO '98, Proceedings. (1998) 370–389
48. Bellare, M., Impagliazzo, R.: A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. IACR Cryptology ePrint Archive **1999** (1999) 24
49. Krawczyk, H.: Lfsr-based hashing and authentication. In: Advances in Cryptology - CRYPTO '94, Proceedings. (1994) 129–139
50. Rogaway, P.: Bucket hashing and its application to fast message authentication. J. Cryptol. **12**(2) (1999) 91–115
51. Moch, A., List, E.: Parallelizable macs based on the sum of prps with security beyond the birthday bound. In: Applied Cryptography and Network Security - ACNS '19, Proceedings. (2019) 131–151
52. Dutta, A., Nandi, M., Talnikar, S.: Beyond birthday bound secure MAC in faulty nonce model. In: Advances in Cryptology - EUROCRYPT '19, Proceedings, Part I. (2019) 437–466
53. Leurent, G., Nandi, M., Sibleyras, F.: Generic attacks against beyond-birthday-bound macs. In: Advances in Cryptology - CRYPTO '18, Proceedings, Part I. (2018) 306–336

54. Aiello, W., Venkatesan, R.: Foiling birthday attacks in length-doubling transformations - benes: A non-reversible alternative to feistel. In: Advances in Cryptology - EUROCRYPT '96, Proceedings. (1996) 307–320
55. Patarin, J., Montreuil, A.: Benes and butterfly schemes revisited. In: Information Security and Cryptology - ICISC '05, Revised Selected Papers. (2005) 92–116
56. Patarin, J.: A proof of security in $o(2^n)$ for the benes scheme. In: Progress in Cryptology - AFRICACRYPT '08, Proceedings. (2008) 209–220
57. Yasuda, K.: A new variant of PMAC: beyond the birthday bound. In: Advances in Cryptology - CRYPTO '11, Proceedings. (2011) 596–609
58. Naito, Y.: Blockcipher-based macs: Beyond the birthday bound without message length. In: Advances in Cryptology - ASIACRYPT '17, Proceedings, Part III. (2017) 446–470
59. Minematsu, K., Iwata, T.: Tweak-length extension for tweakable blockciphers. In: Cryptography and Coding - IMACC '15, Proceedings. (2015) 77–93
60. Kim, S., Lee, B., Lee, J.: Tight security bounds for double-block hash-then-sum macs. In: Advances in Cryptology - EUROCRYPT '20, Proceedings, Part I. (2020) 435–465

Supplementary Material

A Proofs of Proposition 1 and 2

A.1 Proof of Proposition 1

Suppose a compresses to b due to a partition $\mathcal{P}$. Then, we call $\mathcal{P}$ the compressing partition of a and b. For $s \geq 1$, let $p(s)$ denote the claimed statement. We prove the result by induction on s. We first handle the base case, $s = 1$. In this case, we have $b_1 = \sum_{i=1}^{r} a_i$. Thus, $a_i \leq b_1$ for all $i \in [r]$. Now, a term by term comparison gives

$$\prod_{i=1}^{r} (2^n)_{a_i} \geq (2^n)_{b_1},$$

which shows that the base case $p(1)$ is true. Suppose $p(s)$ is true for all $s = n$, for some $n > 1$. We now show that $p(n+1)$ is true.

Let $a = (a_i)_{i\in[r]}$ and $b = (b_j)_{j\in[s+1]}$ be two sequences over $\mathbb{N}$, such that $r \geq s+1$ and a compresses to b. Suppose $\mathcal{P}$ is a compressing partition of a and b. Consider the sequences $a' = (a_i)_{i\in\mathcal{P}_{s+1}}$ and $b' = (b_{s+1})$. We have $|\mathcal{P}_{s+1}| \geq 1$, and $b_{s+1} = \sum_{i\in\mathcal{P}_{s+1}} a_i$, which means a' compresses to b'. Further, $2^n \geq \sum_{i\in\mathcal{P}_{s+1}} a_i$. Thus, we can apply $p(1)$ result on a' and b' to get

$$\prod_{i\in\mathcal{P}_{s+1}} (2^n)_{a_i} \geq (2^n)_{b_{s+1}}. \tag{27}$$

For the remaining, let $a'' = (a_i)_{i\in[r]\setminus\mathcal{P}_{s+1}}$ and $b'' = (b_j)_{j\in[s]}$. Again, we have $r - |\mathcal{P}_{s+1}| \geq s$, and $b_i = \sum_{j\in\mathcal{P}_i} a_j$ for all $i \in [s]$. Thus, we can apply the induction hypothesis for $p(s)$ on a'' and b'' to get

$$\prod_{i\in[r]\setminus\mathcal{P}_{s+1}} (2^n)_{a_i} \geq \prod_{j\in[s]} (2^n)_{b_j}. \tag{28}$$

The combination of Eq. (27) and (28) shows that $p(s+1)$ is true. The result follows by induction. □

A.2 Proof of Proposition 2

For $r \geq 2$, let $p(r)$ denote the claimed statement. We prove the result by induction on r. For now, assume $p(2)$ to be true, as we handle this case later. Suppose the proposition statement, denoted $p(r)$, is true for all $r \geq 2$. We show that the statement $p(r+1)$ is true. Fix some arbitrary $n \in \mathbb{N}$.

Let $a_1, a_2, b_1, b_2, c_1, \ldots, c_{r+1}, d_1, \ldots, d_{r+1} \in \mathbb{N}$, such that $c_i \leq a_i$ and $c_i + d_i \leq a_i + b_j \leq 2^n$, for all $i \in [r+1]$ and $j \in [2]$. Let i' be the smallest index in $[r+1]$, such that $d_{i'} = \min\{d_1, \ldots, d_{r+1}\}$ (such an element exist by well ordering principle). Without loss of generality, we assume that $b_1 \geq b_2$. We compare the terms, $(2^n - c_{i'} - j + 1)$ and $(2^n - a_1 - j + 1)$, for all $j \in [d_{i'}]$. Since $c_{i'} \leq a_1$, we must have $(2^n - c_{i'} - j + 1) \geq (2^n - a_1 - j + 1)$, for all $j \in [d_{i'}]$. Now, we must have $d_{i'} \leq b_1$, otherwise $d_{i'} > b_1 \geq b_2$ which leads to $\sum_{i\in[r]} d_i > b_1 + b_2$. Suppose $d_{i'} < b_1$, then using $(2^n - c_{i'} - j + 1)/(2^n - a_1 - j + 1) \geq 1$, we remove all the $(2^n - c_{i'} - j + 1)$, $(2^n - a_1 - j + 1)$ terms for all $j \in [d_{i'}]$. This reduces the claimed statement to $p(r)$, which is true by hypothesis. If $d_{i'} = b_1$, then we are left with $\prod_{i\in[r+1]\setminus\{i'\}}(2^n - c_i)\cdots(2^n - c_i - d_i + 1)$ on the left, where $r \geq 2$, and $(2^n - a_2)\cdots(2^n - a_2 - b_2 + 1)$ on the right. Using a similar line of argument as above we can again reduce the claimed statement to $p(r)$, which is true by hypothesis. So $p(r+1)$ is true.

Now the base case $p(2)$ can be handled in a similar manner. In this case we assume without loss of generality that $d_1 \leq d_2$ and $b_1 \geq b_2$, where $d_1 + d_2 = b_1 + b_2$. Since $c_1 \leq a_1$, we must have $(2^n - c_1 - j + 1) \geq (2^n - a_1 - j + 1)$, for all $j \in [d_1]$. Now, we must have $d_1 \leq b_1$, otherwise $d_1 > b_1 \geq b_2$ which leads to $d_1 + d_2 > b_1 + b_2$. If $d_1 = b_1$, then after removing all the terms corresponding to (c_1, d_1) and (a_1, b_1), we have $(2^n - c_2)\cdots(2^n - c_2 - d_2 + 1)$ on the left and $(2^n - a_2)\cdots(2^n - a_2 - b_2 + 1)$, where $c_2 \leq a_2$ and $c_2 + b_2 \leq a_2 + b_2$, whence $(2^n - c_2)\cdots(2^n - c_2 - d_2 + 1) \geq (2^n - a_2)\cdots(2^n - a_2 - b_2 + 1)$. If $d_1 < b_1$, then we compare terms from $(2^n - c_2)\cdots(2^n - c_2 - d_2 + 1)$ with $(2^n - a_1 - d_1)\cdots(2^n - a_1 - b_1 + 1)(2^n - a_2)\cdots(2^n - a_2 - b_2 + 1)$. First $(2^n - c_2 - d_2 + j) \geq (2^n - a_2 - b_2 + j)$ for $j \in [b_2]$, as $c_2 + d_2 \leq a_2 + b_2$. We remove all these terms to get $(2^n - c_2)\cdots(2^n - c_2 - d_2 + b_2 + 1)$ on the left and $(2^n - a_1 - d_1)\cdots(2^n - a_1 - b_1 + 1)$ on the right, where the number of terms $d_2 - b_2 = b_1 - d_1$. Since $c_2 \leq a_1$, $(2^n - c_2 - j + 1) \geq (2^n - a_1 - d_1 - j + 1)$ for all $j \in [b_1 - d_1]$. This shows that $p(2)$ is true. □

B Mennink's Attack on CLRW2

In [29] Mennink gave an $O(n^{1/2}2^{3n/4})$ query attack on CLRW2. The attack is generic in nature as it does not exploit the weaknesses in the underlying block cipher. Rather it assumes that the block cipher instances are independent random permutations. Also the attack works for any hash function, including AXU. We briefly describe the attack and refer the readers to [29] for a more concrete and formal description, analysis and experimental verification of the attack.

Attack Description: Suppose in the transcript generated by a distinguisher, there exist four queries (t, m_1, c_1), (t', m_2, c_2), (t, m_3, c_3), and (t', m_4, c_4), such that the following equations hold:

$$\begin{aligned} m_1 \oplus h_1(t) &= m_2 \oplus h_1(t') \\ c_2 \oplus h_2(t') &= c_3 \oplus h_2(t) \\ m_3 \oplus h_1(t) &= m_4 \oplus h_1(t') \end{aligned} \tag{29}$$

Using notations analogous to Figure 3.1, we equivalently have, $x_1 = x_2$; $u_2 = u_3$; and $x_3 = x_4$. Since $x^4 \leftrightsquigarrow y^4$ and $v^4 \leftrightsquigarrow u^4$, looking at the equations generated by the corresponding y and v values, we have $v_1 = y_1 \oplus \lambda(t) = y_2 \oplus \lambda(t) = v_2 \oplus \lambda(t') \oplus \lambda(t) = v_3 \oplus \lambda(t) \oplus \lambda(t') = y_3 \oplus \lambda(t') = v_4$. This immediately gives $u_1 = u_4$, i.e.

$$c_4 \oplus h_2(t') = c_1 \oplus h_2(t). \tag{30}$$

In other words, Eq. (30) is implied by the existence of Eq. (29), and by combining all four equations, we have

$$\begin{aligned} m_1 \oplus m_2 = m_3 \oplus m_4 = \alpha, \\ c_1 \oplus c_4 = c_2 \oplus c_3 = \beta, \end{aligned}$$

where $\alpha = h_1(t) \oplus h_1(t')$ and $\beta = h_2(t) \oplus h_2(t')$. While the distinguisher does not know α and β, it can exploit the relations:

$$m_1 \oplus m_2 = m_3 \oplus m_4, \tag{31}$$

$$c_1 \oplus c_4 = c_2 \oplus c_3. \tag{32}$$

If for some value a we have about 2^n quadruples satisfying

$$m_1 \oplus m_2 = m_3 \oplus m_4 = a, \tag{33}$$

then, for CLRW2, the expected number of solutions for Eq. (31)-(32) is approximately 2 for $a = \alpha$. On the other hand, for $\widetilde{\Pi}$, the expected number of solutions is always close to 1 for any $a \in \{0,1\}^n$. In [29], it has been shown that approximately $2n^{1/2}2^{3n/4}$ queries are sufficient for the distinguisher to ensure that Eq. (33) has about 2^n solutions. Given these many queries the distinguisher can attack by observing the number of solutions for Eq. (31)-(32) for each value of a.

C Proof of Lemma 4.2

Proof. We follow a similar proof approach as considered in Lemma 4.1. We define a binary random vector $\mathsf{I} = (\mathsf{I}_{i,j} : i \neq j)$ where $\mathsf{I}_{i,j}$ takes value 1 if $\mathtt{E}_{i,j}$ holds, otherwise zero. The sample space of the random vector is Ω, the set of all binary vectors indexed by all pairs (i, j). For any vector $w \in \Omega$, we write $\#w$ to represent the number of 1's that appear in w. Let $\Omega_{\leq} = \{w : \#w \leq \frac{1}{\sqrt{\epsilon'}}\}$ and its complement set by $\Omega_{>}$.

We define a random variable $\mathsf{N} = \sum_{i \neq j} \mathsf{I}_{i,j}$, the number of $\mathtt{E}$-events hold. As $\mathtt{E}_{i,j}$ holds with probability at most ϵ,

$$\begin{aligned} q(q-1)\epsilon &\geq \mathsf{Ex}\left[\mathsf{N}\right] \\ &= \sum_{w} \#w \cdot \Pr\left[\mathsf{I} = w\right] \\ &\geq \sum_{w \in \Omega_{\leq}} \#w \cdot \Pr\left[\mathsf{I} = w\right] + \frac{\Pr\left[\mathsf{I} \in \Omega_{>}\right]}{\sqrt{\epsilon'}}. \end{aligned} \tag{34}$$

Let $\mathtt{EEF}$ denote the event that there exists distinct i, j, k, l such that $\mathtt{E}_{i,j} \wedge \mathtt{E}_{k,l} \wedge \mathtt{F}_{i,j,k,l}$. Now we proceed for bounding the probability of the event.

$$\begin{aligned} \Pr\left[\mathtt{EEF}\right] &= \sum_{w} \Pr\left[\mathtt{EEF} \ \wedge \ \mathsf{I} = w\right] \\ &= \sum_{w} \Pr\left[\mathsf{I} = w\right] \times \Pr\left[\mathtt{EEF} \wedge \mathsf{I} = w \mid \mathsf{I} = w\right] \\ &\leq \sum_{w} \Pr\left[\mathsf{I} = w\right] \times \min\{1, (\#w)^2 \cdot \epsilon'\} \\ &= \Pr\left[\mathsf{I} \in \Omega_{>}\right] + \sum_{w \in \Omega_{\leq}} \Pr\left[\mathsf{I} = w\right] \cdot (\#w)^2 \cdot \epsilon' \\ &\leq \Pr\left[\mathsf{I} \in \Omega_{>}\right] + \sum_{w \in \Omega_{\leq}} \Pr\left[\mathsf{I} = w\right] \cdot \#w \cdot \sqrt{\epsilon'} \\ &= \sqrt{\epsilon'} \cdot \Big(\sum_{w \in \Omega_{\leq}} \#w \cdot \Pr\left[\mathsf{I} = w\right] + \frac{\Pr\left[\mathsf{I} \in \Omega_{>}\right]}{\sqrt{\epsilon'}} \Big) \\ &\leq q(q-1)\epsilon \cdot \sqrt{\epsilon'}. \end{aligned}$$

The first inequality follows exactly by the same reason argued in the proof of Lemma 4.1. The last inequality follows from Eq. (34). This completes the proof. □

D Proof of Mirror Theory in Tweakable Settings

The induction is defined on the number of components. Apropos to this, we consider some new parameters. For $i \in [c_1 + c_2 + c_3]$:

– X_i denotes the number of Y-vertices in the previous $i-1$ components.
– U_i denotes the number of V-vertices in the previous $i-1$ components.
– ξ_i denotes the size (number of vertices) of the i-th component. We actually use $\eta_i := \xi_i - 1$ (number of edges in the i-th component).
– for $j \in [\eta_i]$ and $r = \sum_{k=1}^{i-1} \eta_k + j$,
  • $\lambda_j^i := \lambda_r$ (λ value corresponding to the j-th equation of i-th component).
  • $\delta_j^i := \mu(\lambda^{r-1}, \lambda_j^i)$, where $\delta_1^1 = 0$ by convention.
– $\mathfrak{h}_i$ denotes the number of solutions for the sub-system consisting of the first i components of $\mathcal{L}$, denoted $\mathcal{L}_{|i}$. Note that $h_i = \mathfrak{h}_i$ for $i \in [c_1]$, and $h_q = \mathfrak{h}_{c_1+c_2+c_3}$.
– $H_i := \prod_{j\in[\eta_i]} (2^n)_{\mu(\lambda^s,\lambda_j^i)} \cdot \mathfrak{h}_i$, where $s = \sum_{k=1}^{i} \eta_k$.
– $J_i := \begin{cases} (2^n)_{X_i+1}(2^n)_{U_i+1} & i\text{-th component is isolated,} \\ (2^n)_{X_i+1}(2^n)_{U_i+\eta_i} & i\text{-th component is a } \mathcal{Y}\text{-}\star, \\ (2^n)_{X_i+\eta_i}(2^n)_{U_i+1} & i\text{-th component is a } \mathcal{V}\text{-}\star\,. \end{cases}$

Proof Sketch: Inspired by Patarin's mirror theory argument [37,39], we will study the relation between H_i and J_i for all $i \in [c_1 + c_2 + c_3]$. Our goal is to bound $\mathfrak{h}_{c_1+c_2+c_3}$ in terms of $H_{c_1+c_2+c_3}$ and $J_{c_1+c_2+c_3}$. We show that $H_{c_1+c_2+c_3} \geq (1-\epsilon)J_{c_1+c_2+c_3}$, where $\epsilon = O\left(q^2/2^{2n} + \sum_{i=1}^{c_2+c_3} \eta_{c_1+i}^2 q^2/2^{2n}\right)$, which immediately gives the bound for $\mathfrak{h}_{c_1+c_2+c_3}$. This is precisely the motivation behind the definition of H and J.

The proof is given in two steps. First, in section D.1, we bound the number of solutions for the sub-system of equations corresponding to isolated edges, i.e. the first c_1 components. The idea is to apply induction on H_i/J_i for $i \in [c_1]$.

Given the number of solutions for the first c_1 components, we then bound the number of solutions for the remaining c_2+c_3 components (corresponding to star components) in section D.2, which essentially gives a bound for the complete system $\mathcal{L}$. Again, $H_{i'}/J_{i'}$ is analyzed for $i' = c_1 + i$ and $i \in [c_2 + c_3]$. However, we keep the expression in terms of q and η intact.

D.1 Bound for Sub-system Corresponding to Isolated Edges

As noted before, we want to bound $\mathfrak{h}_i$ by induction on i, i.e. we want to evaluate $\mathfrak{h}_{i+1}$ from $\mathfrak{h}_i$. Since isolated components have only one edge, we simply write λ_i and δ_i instead of λ_1^i and δ_1^i. We first give two supplementary results in Lemma D.1 and D.2, which will be used later on to prove the main result.

Lemma D.1. *For $i \in [q_1]$,*

$$\mathfrak{h}_{i+1} = \mathfrak{h}_i\,(2^n - 2i + \delta_{i+1}) + \sum_{(j,k)\in\mathcal{M}} \mathfrak{h}'_i(j,k,\lambda_{i+1}),$$

where

$$\mathcal{M} = \{(j,k) : j,k \in [i], j \neq k, \lambda_{i+1} \neq \lambda_j, \lambda_{i+1} \neq \lambda_k\},$$

and $\mathfrak{h}'_i(j,k,\lambda_{i+1})$ denotes the number of solutions of $\mathcal{L}'_{|i}(j,k,\lambda_{i+1}) := \mathcal{L}_{|i} \cup \{Y_j \oplus V_k = \lambda_{i+1}\}$, for some $j,k \in [i]$.

Proof. Let $\mathcal{S}_i$ denote the solution space of $\mathcal{L}_{|i}$, i.e. $\mathfrak{h}_i = |\mathcal{S}_i|$. For a fix $(y^i, v^i) \in \mathcal{S}_i$, we want to compute the number of (y_{i+1}, v_{i+1}) pairs such that $(y^{i+1}, v^{i+1}) \in \mathcal{S}_{i+1}$. Now, some pair $(x, x \oplus \lambda_{i+1})$ is valid if $x \neq y_j$ and $x \oplus \lambda_{i+1} \neq v_k$, for $j, k \in [i]$. This means that $x \notin \mathcal{Y} \cup \mathcal{V}$, where $\mathcal{Y} = \{y_j : j \in [i]\}$ and $\mathcal{V} = \{v_j \oplus \lambda_{i+1} : j \in [i]\}$. As all y_j values are pairwise distinct and v_j values are pairwise distinct, we must have $|\mathcal{Y}| = |\mathcal{V}| = i$. Thus, we have

$$\begin{aligned}
\mathfrak{h}_{i+1} &= \sum_{(y^i,v^i)\in\mathcal{S}_i} (2^n - |\mathcal{Y} \cup \mathcal{V}|) \\
&= \sum_{(y^i,v^i)\in\mathcal{S}_i} (2^n - |\mathcal{Y}| - |\mathcal{V}| + |\mathcal{Y} \cap \mathcal{V}|) \\
&= \mathfrak{h}_i \cdot (2^n - 2i) + \sum_{(y^i,v^i)\in\mathcal{S}_i} |\mathcal{Y} \cap \mathcal{V}| \\
&= \mathfrak{h}_i \cdot (2^n - 2i) + \sum_{(y^i,v^i)\in\mathcal{S}_i} \sum_{j,k\in[i]} \phi(j,k) \\
&\overset{1}{=} \mathfrak{h}_i \cdot (2^n - 2i) + \sum_{j,k\in[i]} \mathfrak{h}'_i(j,k,\lambda_{i+1}) \\
&\overset{2}{=} \mathfrak{h}_i \cdot (2^n - 2i) + \mathfrak{h}_i \cdot \delta_{i+1} + \sum_{(j,k)\in\mathcal{M}} \mathfrak{h}'_i(j,k,\lambda_{i+1}) \\
&= \mathfrak{h}_i \cdot (2^n - 2i + \delta_{i+1}) + \sum_{(j,k)\in\mathcal{M}} \mathfrak{h}'_i(j,k,\lambda_{i+1}), \qquad (35)
\end{aligned}$$

where $\phi(j,k)$ is the indicator variable that takes the value of 1 when $y_j \oplus v_k = \lambda_{i+1}$, and 0 otherwise. The equality 1 follows from the definition of $\mathfrak{h}'_i(j,k,\lambda_{i+1})$, and the equality 2 follows from the fact that exactly δ_{i+1} (j,k) pairs exist such that $k = j$, $\lambda_{i+1} = \lambda_j$, and $y_j \oplus v_j = \lambda_{i+1}$. For these the number of solutions is exactly the same as $\mathfrak{h}_i$ (since $Y_j \oplus V_k = \lambda_{i+1}$ is already in $\mathcal{L}_{|i}$). The remaining valid (j,k) pairs, must have $\lambda_j, \lambda_k \neq \lambda_{i+1}$, else they contradict $\mathcal{L}$. The set of these remaining (j,k) pairs is the set $\mathcal{M}$. □

The following corollary of Lemma D.1 will be quite useful. The proof is immediate from the proof of Lemma D.1.

Corollary D.1. *For $i \geq 1$, let $\widehat{\mathcal{L}}_{i+1}$ be a system of $i+1$ equations such that $\xi_{\max}(\widehat{\mathcal{L}}_{i+1}) = 2$. Then, for any sub-system $\widehat{\mathcal{L}}_i$ consisting of i equations from $\widehat{\mathcal{L}}_{i+1}$, we have*

$$(2^n - 2i)\widehat{\mathfrak{h}}_i \leq \widehat{\mathfrak{h}}_{i+1} \leq (2^n - i)\widehat{\mathfrak{h}}_i,$$

where $\widehat{\mathfrak{h}}_i$ and $\widehat{\mathfrak{h}}_{i+1}$ denote the number of solutions of $\widehat{\mathcal{L}}_i$ and $\widehat{\mathcal{L}}_{i+1}$, respectively.

Lemma D.2. *For all $(j,k) \in \mathcal{M}$, and for all $\beta \in \{0,1\}^n$,*

$$\mathfrak{h}'_i(j,k,\beta) \geq \frac{\mathfrak{h}_i}{2^n - i + 1} \cdot \left(1 - \frac{2(i-2)}{2^n - 2(i-2)}\right).$$

Proof. We are interested in $\mathfrak{h}'_i(j,k,\beta)$, which is the number of solutions of $\mathcal{L}'_{|i}(j,k,\beta)$, $j,k \in \mathcal{M}$. The sub-system containing j and k equations is of the form

$$Y_j \oplus V_j = \lambda_j, \quad Y_j \oplus V_k = \beta, \quad Y_k \oplus V_k = \lambda_k,$$

where once we fix $Y_j = y_j$, all other unknowns are completely determined by linearity. Thus, $\mathfrak{h}'_i(j,k,\beta)$ is at most $\widehat{\mathfrak{h}}_{i-1}$, where $\widehat{\mathfrak{h}}_{i-1}$ is the number of solutions of $\widehat{\mathcal{L}}_{|i-1} := \mathcal{L}'_{|i}(j,k,\beta) \setminus \{Y_j \oplus V_k = \beta, Y_k \oplus V_k = \lambda_k\}$, the system obtained by removing the equations $Y_j \oplus V_k = \beta$ and $Y_k \oplus V_k = \lambda_k$ from $\mathcal{L}'_{|i}(j,k,\beta)$. Now a solution among the $\widehat{\mathfrak{h}}_{i-1}$ solutions of $\widehat{\mathcal{L}}_{|i-1}$ is not valid to be counted in $\mathfrak{h}'_i(j,k,\beta)$, if there exists $\ell \in [i] \setminus \{k\}$, such that $y_j \oplus v_\ell = \beta$ or $y_j \oplus v_\ell = \beta \oplus \lambda_k \oplus \lambda_\ell$. The first case leads to $V_k = V_\ell$, and the second case leads to $Y_k = Y_\ell$, where $k \neq \ell$ is obvious. Let $\widehat{\mathcal{L}}'_{|i-1}(j,\ell,\beta) := \widehat{\mathcal{L}}_{|i-1} \cup \{Y_j \oplus V_\ell = \beta\}$ and $\widehat{h}'_{i-1}(j,\ell,\beta)$ be the number of solutions $\widehat{\mathcal{L}}'_{|i-1}(j,\ell,\beta)$. Therefore, the two cases correspond to the terms $\widehat{\mathfrak{h}}'_{i-1}(j,\ell,\beta)$ and $\widehat{\mathfrak{h}}'_{i-1}(j,\ell',\beta \oplus \lambda_k \oplus \lambda_{\ell'})$, whence we have

$$\mathfrak{h}'_i(j,k,\beta) \geq \widehat{\mathfrak{h}}_{i-1} - \sum_{\ell \in [i]\setminus\{j,k\}} \widehat{\mathfrak{h}}'_{i-1}(j,\ell,\beta) - \sum_{\ell' \in [i]\setminus\{j,k\}} \widehat{\mathfrak{h}}'_{i-1}(j,\ell',\beta \oplus \lambda_k \oplus \lambda_{\ell'})$$

Let $\widehat{\mathcal{L}}_{|i-2,\ell} := \widehat{\mathcal{L}}'_{|i-1}(j,\ell,\beta) \setminus \{Y_j \oplus V_\ell = \beta, Y_\ell \oplus V_\ell = \lambda_\ell\}$ and $\widehat{\mathcal{L}}_{|i-2,\ell'} := \widehat{\mathcal{L}}'_{|i-1}(j,\ell',\beta \oplus \lambda_k \oplus \lambda_{\ell'}) \setminus \{Y_j \oplus V_{\ell'} = \beta \oplus \lambda_k \oplus \lambda_{\ell'}, Y_{\ell'} \oplus V_{\ell'} = \lambda_{\ell'}\}$. Let $\widehat{\mathfrak{h}}_{i-2,\ell}$ and $\widehat{\mathfrak{h}}_{i-2,\ell'}$ be the number of solutions for $\widehat{\mathcal{L}}_{|i-2,\ell}$ and $\widehat{\mathcal{L}}_{|i-2,\ell'}$. Using similar line of argument as above we bound $\widehat{\mathfrak{h}}'_{i-1}(j,\ell,\beta) \leq \widehat{\mathfrak{h}}_{i-2,\ell}$ and $\widehat{\mathfrak{h}}'_{i-1}(j,\ell',\beta \oplus \lambda_k \oplus \lambda_{\ell'}) \leq \widehat{\mathfrak{h}}_{i-2,\ell'}$. Finally, we have

$$\begin{aligned}
\mathfrak{h}'_i(j,k,\beta) &\geq \widehat{\mathfrak{h}}_{i-1} - \sum_{\ell \in [i]\setminus\{j,k\}} \widehat{\mathfrak{h}}_{i-2,\ell} - \sum_{\ell' \in [i]\setminus\{j,k\}} \widehat{\mathfrak{h}}_{i-2,\ell'} \\
&\geq \widehat{\mathfrak{h}}_{i-1} - (i-2)\widehat{\mathfrak{h}}_{i-2,\ell} - (i-2)\widehat{\mathfrak{h}}_{i-2,\ell'} \\
&\overset{1}{\geq} \widehat{\mathfrak{h}}_{i-1}\left(1 - \frac{2(i-2)}{2^n - 2(i-2)}\right) \\
&\overset{2}{\geq} \frac{\mathfrak{h}_i}{2^n - i + 1}\left(1 - \frac{2(i-2)}{2^n - 2(i-2)}\right),
\end{aligned}$$

where inequalities 1 and 2 follow from Corollary D.1. Note that, we switch from $\widehat{\mathfrak{h}}_{i-2,\ell}$ and $\widehat{\mathfrak{h}}_{i-2,\ell'}$ to $\widehat{\mathfrak{h}}_{i-1}$ by reintroducing the equation $Y_\ell \oplus V_\ell = \lambda_\ell$ and $Y_{\ell'} \oplus V_{\ell'} = \lambda_{\ell'}$, respectively, and from $\widehat{\mathfrak{h}}_{i-1}$ to $\mathfrak{h}_i$ by reintroducing the equation $Y_k \oplus V_k = \lambda_k$. The readers may use Figure D.1 to get a pictorial view of the switchings between different systems of equations. □

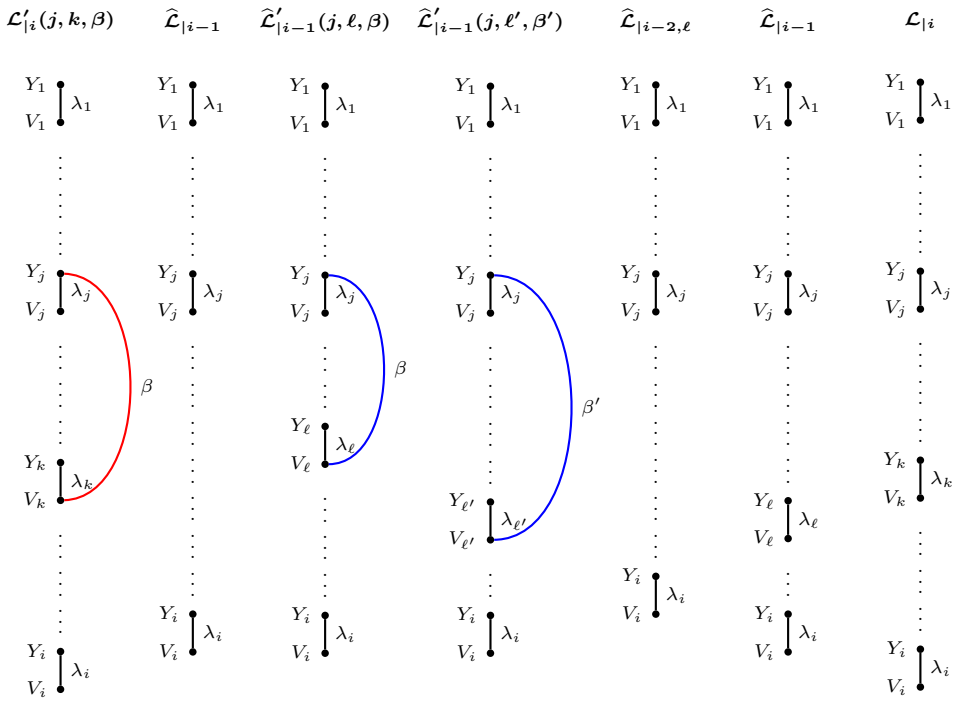

Fig. D.1: The switchings used in the proof of Lemma D.2. From left to right: $\mathcal{L}'_{|i}(j,k,\beta)$ is the system $\mathcal{L}_{|i} \cup \{Y_j \oplus V_k = \beta\}$; $\widehat{\mathcal{L}}_{|i-1}$ is obtained by removing the equations involving V_k from $\mathcal{L}'_{|i}(j,k,\beta)$; $\widehat{\mathcal{L}}'_{|i-1}(j,\ell,\beta)$ is the system $\widehat{\mathcal{L}}_{|i-1} \cup \{Y_j \oplus V_\ell = \beta\}$; $\widehat{\mathcal{L}}'_{|i-1}(j,\ell',\beta')$ is the system $\widehat{\mathcal{L}}_{|i-1} \cup \{Y_j \oplus V_{\ell'} = \beta'\}$, where $\beta' = \beta \oplus \lambda_k \oplus \lambda_{\ell'}$; $\widehat{\mathcal{L}}_{|i-2,\ell}$ is obtained by removing the equations involving V_ℓ from $\widehat{\mathcal{L}}'_{|i-1}(j,\ell,\beta)$. Note that, there should have been two $\widehat{\mathcal{L}}_{|i-2}$ switchings, one each for $\widehat{\mathcal{L}}'_{|i-1}(j,\ell,\beta)$ and $\widehat{\mathcal{L}}'_{|i-1}(j,\ell',\beta')$. We have drawn just once for economical reasons. Similar clarification applies to switchings from $\widehat{\mathcal{L}}_{|i-2}$ to $\widehat{\mathcal{L}}_{|i-1}$ (we only show for ℓ).

Remark 1. In [37, Theorem 11] a result similar to Lemma D.2 has been proved for random function scenario. While the proof of that theorem is correct, there is a notational issue which is worth pointing out. The $\mathfrak{h}'$ notation is used in an unparametrized fashion, with an explicit hint in [37, Theorem 8] that this is done for simplification. But this simplification leads to a rather peculiar technical issue in [37, Theorem 11], where both lower and upper bounds are required on $\mathfrak{h}'$ values, requiring different switchings. Without the parametrization it is difficult to understand (and verify) the switchings.

Remark 2. The proof of Lemma D.2 should also give an idea of the proof complexity. Since we only want $\epsilon = O(q^4/2^{3n})$, we needed a somewhat crude estimate of $\mathfrak{h}'$ values. In actual mirror theory as we move towards $\epsilon = O(q/2^n)$, we have to make a good estimate of $\mathfrak{h}'$ values, which does not seem easy.

Now, we state the main result of this section.

Lemma D.3. *For $q_1 < 2^{n-2}$, we have*

$$\frac{H_{q_1}}{J_{q_1}} \geq \left(1 - \frac{13q_1^4}{2^{3n}} - \frac{2q_1^2}{2^{2n}}\right).$$

Proof. We prove by induction on $i \in [q_1]$, the number of components. First, $H_1 = 2^{2n} = J_1$. So the statement is true for $i = 1$. By definition, the ratio $\frac{H_{i+1}}{H_i} = (2^n - \delta_{i+1}) \cdot \frac{\mathfrak{h}_{i+1}}{\mathfrak{h}_i}$, and $J_{i+1} = (2^n - i)^2 J_i$. So we have

$$\frac{H_{i+1}}{J_{i+1}} = \frac{(2^n - \delta_{i+1})\frac{\mathfrak{h}_{i+1}}{\mathfrak{h}_i}}{(2^n - i)^2} \frac{H_i}{J_i}. \tag{36}$$

From Lemma D.1 and D.2, we have

$$\mathfrak{h}_{i+1} \geq \mathfrak{h}_i \left((2^n - 2i + \delta_{i+1}) + \frac{|\mathcal{M}|}{2^n - i + 1}\left(1 - \frac{2(i-2)}{2^n - 2(i-2)}\right)\right). \tag{37}$$

Recall that $\mathcal{M} = \{(j,k) : j,k \in [i], j \neq k, \lambda_j, \lambda_k \neq \lambda_{i+1}\}$. As there are δ_{i+1} $i' \in [i]$ such that $\lambda_{i+1} = \lambda_{i'}$, we must have $|\mathcal{M}| \geq (i - \delta_{i+1})(i - \delta_{i+1} - 1)$. On substituting this value for $|\mathcal{M}|$ in Eq. (37), and using the resulting lower bound for $\mathfrak{h}_{i+1}$ in Eq. (36), we get

$$\frac{H_{i+1}}{J_{i+1}} \geq \boxed{\frac{(2^n - \delta_{i+1})\left((2^n - 2i + \delta_{i+1}) + \frac{(i-\delta_{i+1})(i-\delta_{i+1}-1)}{2^n - i + 1}\left(1 - \frac{2(i-2)}{2^n - 2(i-2)}\right)\right)}{(2^n - i)^2}} \frac{H_i}{J_i}.$$

Let the boxed expression be A. We first simplify this term.

$$\begin{aligned}
A &\geq \frac{(2^n - \delta_{i+1})\left((2^n - 2i + \delta_{i+1}) + \frac{(i-\delta_{i+1})(i-\delta_{i+1}-1)}{2^n - i + 1}\left(1 - \frac{2(i-2)}{2^n - 2(i-2)}\right)\right)}{(2^n - i)^2} \\
&\overset{1}{\geq} \frac{(2^n - \delta_{i+1})(2^n - 2i + \delta_{i+1}) + \frac{(2^n - \delta_{i+1})(i-\delta_{i+1})(i-\delta_{i+1}-1)}{2^n} - \frac{16}{3}\frac{i^3}{2^n}}{(2^n - i)^2} \\
&\overset{2}{\geq} 1 - \frac{(i - \delta_{i+1}) + \frac{(i-\delta_{i+1})^2\delta_{i+1}}{2^n} - \frac{(i-\delta_{i+1})\delta_{i+1}}{2^n} + \frac{16}{3}\frac{i^3}{2^n}}{(2^n - i)^2} \\
&\overset{3}{\geq} 1 - \frac{13i^3}{2^{3n}} - \frac{2i}{2^{2n}}.
\end{aligned}$$

At inequality 1, we use $i \leq q_1 \leq 2^{n-2}$, $(i-2), (i-\delta_{i+1}) < i$, and $(2^n - \delta_{i+1}), (2^n - i + 1) < 2^n$; inequality 2 is just a simplification; and at inequality 3, we use $(i - \delta_{i+1}), \delta_{i+1} \leq i$ and $(2^n - i)^2 \leq 2^{n-1}$. Now, we have

$$\frac{H_{i+1}}{J_{i+1}} \geq \left(1 - \frac{13i^3}{2^{3n}} - \frac{2i}{2^{2n}}\right) \times \frac{H_i}{J_i}$$

$$\overset{1}{\geq} \left(1 - \frac{13i^3}{2^{3n}} - \frac{2i}{2^{2n}}\right)^i$$
$$\geq \left(1 - \frac{13i^4}{2^{3n}} - \frac{2i^2}{2^{2n}}\right).$$

Inequality 1 follows from recursive application of the induction hypothesis. The result follows by induction. □

D.2 Bound for Sub-system Corresponding to Star Components

At this point, we have the bound for the sub-system corresponding to the q_1 isolated edges, and we want to extend it to get the bound on $\mathfrak{h}_{q_1+c_2+c_3}$. For simplicity we let $i' = q_1 + i = c_1 + i$. Thus, $c_1 + c_2 + c_3 = (c_2 + c_3)'$. We follow exactly the same approach as before in case of isolated edges.

For $i' - 1 \geq 0$, we analyze the ratio $\frac{H_{i'}}{J_{i'}}$. Note that $J_{i'}$ depends on the type of i'-th component ($\mathcal{Y}$-$\star$ or $\mathcal{V}$-$\star$). However, it can be easily seen that the two expressions are symmetric. Without loss of generality, we assume that the i'-th component is $\mathcal{Y}$-$\star$. Then, we have

$$\frac{H_{i'}}{J_{i'}} = \boxed{\frac{\prod_{j=1}^{\eta_{i'}}(2^n - \delta_j^{i'})^{\frac{\mathfrak{h}_{i'}}{\mathfrak{h}_{i'-1}}}}{(2^n - X_{i'})(2^n - U_{i'})_{\eta_{i'}}}} \times \frac{H_{i'-1}}{J_{i'-1}}.$$

Let the boxed expression be A. We first simplify this term. In Lemma D.5, we show that

$$\frac{\mathfrak{h}_{i'}}{\mathfrak{h}_{i'-1}} \geq \left(2^n - X_{i'} - \eta_{i'}U_{i'} + \sum_{j=1}^{\eta_{i'}} \delta_j^{i'}\right).$$

Thus, we have

$$A \geq \frac{\prod_{j=1}^{\eta_{i'}}(2^n - \delta_j^{i'})(2^n - X_{i'} - \eta_{i'}U_{i'} + \sum_{k=1}^{\eta_{i'}} \delta_k^{i'})}{(2^n - X_{i'})(2^n - U_{i'})_{\eta_{i'}}}$$
$$\geq 1 - \frac{\overbrace{(2^n - X_{i'})(2^n - U_{i'})_{\eta_{i'}}}^{B} - \overbrace{\prod_{j=1}^{\eta_{i'}}(2^n - \delta_j^{i'})(2^n - X_{i'} - \eta_{i'}U_{i'} + \sum_{k=1}^{\eta_{i'}} \delta_k^{i'})}^{C}}{(2^n - X_{i'})(2^n - U_{i'})_{\eta_{i'}}}. \tag{38}$$

We need both lower and upper bounds on B. Using the facts that $X_{i'}, U_{i'} + \eta_{i'} < q$, and $\xi_{\max} q < 2^{n-1}$, we get $B \geq 2^{n(\eta_{i'}+1)-1}$. Now, we derive an upper bound on B.

$$B = (2^n - X_{i'})(2^n - U_{i'})_{\eta_{i'}}$$
$$\leq (2^n - X_{i'})(2^n - U_{i'})^{\eta_{i'}}$$

$$\leq (2^n - X_{i'}) \left(2^{n\eta_{i'}} - \eta_{i'} U_{i'} 2^{n(\eta_{i'}-1)} + \eta_{i'}^2 U_{i'}^2 2^{n(\eta_{i'}-2)}\right)$$
$$\leq 2^{n(\eta_{i'}+1)} - \eta_{i'} U_{i'} 2^{n\eta_{i'}} + \eta_{i'}^2 U_{i'}^2 2^{n(\eta_{i'}-1)} - X_{i'} 2^{n\eta_{i'}} + \eta_{i'} X_{i'} U_{i'} 2^{n(\eta_{i'}-1)}. \tag{39}$$

We also need a lower bound on C.

$$\begin{aligned} C &= \prod_{j=1}^{\eta_{i'}} \left(2^n - \delta_j^{i'}\right) \left(2^n - X_{i'} - \eta_{i'} U_{i'} + \sum_{k=1}^{\eta_{i'}} \delta_k^{i'}\right) \\ &\geq \left(2^{n\eta_{i'}} - \sum_{j=1}^{\eta_{i'}} \delta_j^{i'} 2^{n(\eta_{i'}-1)}\right) \left(2^n - X_{i'} - \eta_{i'} U_{i'} + \sum_{k=1}^{\eta_{i'}} \delta_k^{i'}\right) \\ &\geq 2^{n(\eta_{i'}+1)} - X_{i'} 2^{n\eta_{i'}} - \eta_{i'} U_{i'} 2^{n\eta_{i'}} - \left(\sum_{j=1}^{\eta_{i'}} \delta_j^{i'}\right)^2 2^{n(\eta_{i'}-1)}. \end{aligned} \tag{40}$$

On substituting the bounds of B and C in Eq. (38), we get

$$\begin{aligned} A &\geq \frac{\eta_{i'}^2 U_{i'}^2 2^{n(\eta_{i'}-1)} + \eta_{i'} X_{i'} U_{i'} 2^{n(\eta_{i'}-1)} + \left(\sum_{j=1}^{\eta_{i'}} \delta_j^{i'}\right)^2 2^{n(\eta_{i'}-1)}}{2^{n(\eta_{i'}+1)-1}} \\ &\overset{1}{\geq} \frac{\eta_{i'}^2 q^2 2^{n(\eta_{i'}-1)} + \eta_{i'} q^2 2^{n(\eta_{i'}-1)} + q^2 2^{n(\eta_{i'}-1)}}{2^{n(\eta_{i'}+1)-1}} \\ &\overset{2}{\geq} \frac{4\eta_{i'}^2 q^2}{2^{2n}}. \end{aligned} \tag{41}$$

At inequality 1, we use the fact that $X_{i'}, Y_{i'} \leq q$ and $\sum_{j=1}^{\eta_{i'}} \delta_j^{i'} < q$ ($\lambda_j^{i'}$ can occur at most once in any component). At inequality 2, we use the fact that $\eta_{i'}^2 > \eta_{i'} + 1$ as $\eta_{i'} > 2$. Therefore, we have

$$\frac{H_{i'}}{J_{i'}} \geq \left(1 - \frac{4\eta_{i'}^2 q^2}{2^{2n}}\right) \times \frac{H_{i'-1}}{J_{i'-1}}.$$

In combination with Lemma D.3, this immediately gives the bound on $\frac{H_{c_1+c_2+c_3}}{J_{c_1+c_2+c_3}}$ in Lemma D.4.

Lemma D.4. *For $q \leq 2^{n-2}$ and $\xi_{\max} \leq 2^n/2q$, we have*

$$\frac{H_{c_1+c_2+c_3}}{J_{c_1+c_2+c_3}} \geq \left(1 - \frac{13q^4}{2^{3n}} - \frac{2q^2}{2^{2n}} - \left(\sum_{i=1}^{c_2+c_3} \eta_{c_1+i}^2\right) \frac{4q^2}{2^{2n}}\right).$$

Theorem 5.1 follows from the definition of H, J and Lemma D.4.

Lemma D.5. $\mathfrak{h}_{i'} \geq \left(2^n - X_{i'} - \eta_{i'} U_{i'} + \sum_{j=1}^{\eta_{i'}} \delta_j^{i'}\right) \cdot \mathfrak{h}_{i'-1}.$

Proof. Let $\mathcal{S}_{i'-1}$ denote the solution space of $\mathcal{L}_{|i'-1}$. Let $r = \sum_{j=1}^{i'-1} \eta_j$. For a fixed $(y^r, v^r) \in \mathcal{S}_{i'-1}$, we want to compute the number of solutions for $\mathcal{L}_{|i'}$. Since, this is a $\mathcal{Y}$-$\star$ component, it is sufficient to choose an assignment for $Y_{i'}$ (center of the i'-th component) value and $V_j^{i'} = Y_{i'} \oplus \lambda_j^{i'}$. Now, an assignment x is invalid if $x \in \mathcal{Y} \cup \mathcal{V}$, where $\mathcal{Y} = \{y_j : j \in [r]\}$ and $\mathcal{V} = \{v_j \oplus \lambda_k^{i'} : j \in [r], k \in [\eta_{i'}]\}$. Clearly, $|\mathcal{Y}| = X_{i'}$ and $|\mathcal{V}| \leq \eta_{i'} U_{i'}$. Further, exactly $\sum_{j=1}^{\eta_{i'}} \delta_j^{i'}$ previous equations share λ value with some equation in the i'-th component, whence $|\mathcal{Y} \cap \mathcal{V}| \geq \sum_{j=1}^{\eta_{i'}} \delta_j^{i'}$. Thus, we have

$$\begin{aligned}
\mathfrak{h}_{i'} &= \sum_{(y^r, v^r) \in \mathcal{S}_{i'}} (2^n - |\mathcal{Y} \cup \mathcal{V}|) \\
&= \sum_{(y^r, v^r) \in \mathcal{S}_{i'}} (2^n - |\mathcal{Y}| - |\mathcal{V}| + |\mathcal{Y} \cap \mathcal{V}|) \\
&\geq \sum_{(y^r, v^r) \in \mathcal{S}_{i'}} \left(2^n - X_{i'} - \eta_{i'} U_{i'} + \sum_{j=1}^{\eta_{i'}} \delta_j^{i'} \right) \\
&= \left(2^n - X_{i'} - \eta_{i'} U_{i'} + \sum_{j=1}^{\eta_{i'}} \delta_j^{i'} \right) \cdot \mathfrak{h}_{i'-1}.
\end{aligned}$$

□

E Proof of Lemma 6.1

Property 1 holds by definition and the non-existence of bad hash key condition 1. Property 2 holds due to the non-existence of bad hash key conditions 2 and 3. Property 3 holds due to the non-existence of bad hash key conditions 4, 5, 6, and 7. Property 4 holds due to non-existence of bad hash key conditions 4 and 5. It is easy to verify that given Property 1, 2, 3, and 4, Figure 6.1 enumerates all possible types of components of $\mathcal{G}$. □

F Proof of Security of DbHtS-p

The analysis of DbHtS-p would be similar to the analysis of CLRW2 presented in this paper. The variables arising in DbHtS-p computation is analogously notated as in CLRW2 (see Figure F.1). Specifically, we have the following connection between the notations for DbHtS-p and CLRW2:

- x^q and u^q in DbHtS-p corresponds to x^q and u^q in CLRW2. Here, $x^q = h_1(m^q)$ and $u^q = h_2(m^q)$.
- y^q and v^q in DbHtS-p corresponds to y^q and v^q in CLRW2.
- Similar to CLRW2, in DbHtS-p $x^q \leftrightsquigarrow y^q$ and $u^q \leftrightsquigarrow v^q$. Note that, in DbHtS-p $v^q = E_{k_2}(u^q)$, whereas in CLRW2 $u^q = E_{k_2}(v^q)$. However, this does not affect the permutation compatibility property.

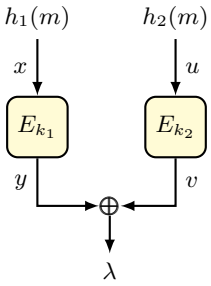

Fig. F.1: The DbHtS-p construction.

– λ^q in DbHtS-p corresponds to λ^q in CLRW2. Therefore, $v^q \oplus y^q = \lambda^q$.

Initial Setup: The first step of replacing the block cipher instantiations with independent uniform random permutations Π_1 and Π_2 incurs a cost of $2\mathbf{Adv}^{\mathsf{prp}}_E(q, t')$. For the sake of simplicity, we call the resulting construction DbHtS-p.

Oracle Description and Sampling Mechanism: The real and ideal oracles can be described in a similar manner as in case of CLRW2, except a small change. For all $i \in [q]$, $\lambda_i \xleftarrow{\$} \{0,1\}^n$ in the ideal world, and $\lambda_i = \mathsf{DbHtS\text{-}p}(m_i)$ in the real world.

Definition of Bad Transcript and its Analysis: We again use the same set of bad transcripts and bound the probability of realizing a bad transcript, denoted ϵ_{bad}, as

$$\epsilon_{\mathsf{bad}} \leq q^2\epsilon^2 + \frac{q^2\epsilon}{2^n} + 2q^2\epsilon^{1.5} + \frac{16q^4\epsilon}{2^{2n}} + \frac{9q^4\epsilon^2}{2^n}. \tag{42}$$

Here the only notable difference is the bound on $\Pr[\mathtt{H}_2]$ and $\Pr[\mathtt{H}_3]$. Since, now the λ values are uniform at random, $\Pr[\mathtt{H}_2] \leq \binom{q}{2}\epsilon 2^{-n}$ and $\Pr[\mathtt{H}_3] \leq \binom{q}{2}\epsilon 2^{-n}$. All other bad events are bounded identically to the bad events in case of CLRW2.

Good Transcript Analysis: For a fixed good transcript ω, in the real world the interpolation probability is bounded as in case of CLRW2, i.e.

$$\Pr[\Theta_1 = \omega] = \frac{1}{|\mathcal{H}|^2} \times \frac{1}{(2^n)_{q_1+c_2+q_3+2c_4+q_5-c_5}} \times \frac{1}{(2^n)_{q_1+q_2+c_3+q_4-c_4+2c_5}}. \tag{43}$$

In the ideal world, using Corollary 5.1 we get

$$\Pr[\Theta_0 = \omega] \leq \frac{1}{|\mathcal{H}|^2} \times \frac{1}{\zeta(\omega)} \times \frac{2^{n(q_1+q_2+q_3)}}{2^{nq}(2^n)_{p_1}(2^n)_{p_2}(2^n)^{c_4+c_5}}, \tag{44}$$

where $p_1 = q_1 + c_2 + q_3$, $p_2 = q_1 + q_2 + c_3$, and

$$\zeta(\omega) = \left(1 - \frac{13q^4}{2^{3n}} - \frac{2q^2}{2^{2n}} - \left(\sum_{i=1}^{c_2+c_3} \eta^2_{c_1+i}\right)\frac{4q^2}{2^{2n}}\right).$$

On dividing Eq. (43) by (44) and doing some simplification, we get

$$\frac{\Pr\left[\Theta_1 = \omega\right]}{\Pr\left[\Theta_0 = \omega\right]} \geq \zeta(\omega).$$

Using Lemma 2.1, we get

$$\begin{aligned}\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{DbHtS\text{-}p}}(q) &\leq \frac{2q^2}{2^{2n}} + \frac{13q^4}{2^{3n}} + \frac{4q^2}{2^{2n}}\mathsf{Ex}\left[\sum_{i=1}^{c_2+c_3}\eta^2_{c_1+i}\right] + \epsilon_{\mathsf{bad}} \\ &\leq \frac{2q^2}{2^{2n}} + \frac{13q^4}{2^{3n}} + \frac{16q^4\epsilon}{2^{2n}} + \epsilon_{\mathsf{bad}}. \end{aligned} \tag{45}$$

The result follows from Eq. (42) and (45). □

Note that, the application of alternating events/collisions lemma (or a similar result) seems indispensable, even if one assumes that the fundamental theorem of mirror theory holds.