Fine-Tuning Ideal Worlds for the Xor of Two Permutation Outputs

Wonseok Choi[1], Minki Hhan[2], Yu Wei[1], and Vassilis Zikas[1]

[1] Purdue University, West Lafayette, IN, USA
{wonseok, yuwei}@purdue.edu, {vzikas}@cs.purdue.edu
[2] KIAS, Seoul, Korea
{minkihhan}@kias.re.kr

Abstract. Security proofs of symmetric-key primitives typically consider an idealized world with access to a (uniformly) random function. The starting point of our work is the observation that such an ideal world leads to underestimating the actual security of certain primitives. As a demonstrating example, XoP2, which relies on two independent random permutations, is proven to exhibit far superior concrete security compared to XoP, which employs a single permutation with domain separation. But the main reason for this is an artifact of the idealized model used in the proof, in particular, that (in the random-function-ideal world) XoP might hit a trivially bad event (outputting $\mathbf{0}$) which does not occur in the real/domain-separated world.

Motivated by this, we put forth the analysis of such primitives in an updated ideal world, which we call the *fine-tuned* setting, where the above artifact is eliminated. We provide fine-tuned (and enhanced) security analyses for XoP and XoP-based MACs: nEHtM and DbHtS. Our analyses demonstrate that the security of XoP-based and XoP2-based constructions are, in fact, far more similar than what was previously proven. Concretely, for the number of users u and the maximum number of queries per user q_m, we show that the multi-user "fine-tuned" security bound of XoP can be proven as $O\left(u^{0.5}{q_m}^2/2^{2n}\right)$ via the Squared-ratio method proposed by Chen et al. [CRYPTO'23], resulted to the same security bound of XoP2 proven there. We also show the compatibility of the fine-tuned model with the Chi-squared method proposed by Dai et al. [CRYPTO'17], and show that XoP and XoP2 enjoy the same security bound in the fine-tuned setting regardless of proving tools.

Finally, we turn to the security analysis of MACs in the multi-user setting, where the effect of transitioning the proofs to the fine-tuned setting is even higher. Concretely, we are able to prove unexpected improvements in the security bounds for both nEHtM and DbHtS. Our security proofs rely on a fine-tuned and extended version of Mirror theory for both lower and upper bounds, which yields more versatile and improved security proofs. Of independent interest, this extension allows us to prove the multi-user MAC security of nEHtM in the nonce-misuse model, while the previous analysis only applied to the multi-user PRF security in the nonce-respecting model. As a side note, we also point out (and fix) a flaw in the original analysis of Chen et al..

Table of Contents

1 Introduction

Block ciphers are often regarded as pseudorandom permutations (PRPs) in cryptography. This so-called standard model in symmetric cryptography assumes that distinguishing a secure block cipher from a random permutation is almost impossible until a certain number of encryption and decryption queries have been made, enabling a black-box methodology for block ciphers. On the other hand, many symmetric-key constructions, such as message authentication codes or authenticated encryptions, rely on pseudorandom functions (PRFs) as foundational building blocks to achieve beyond-birthday-bound security [2, 3, 7, 15]. However, substituting PRFs with PRPs in such constructions results in compromised security, leading to vulnerabilities concerning the birthday-bound [4, 5, 6, 9, 26, 27]. Because block ciphers take many advantages, e.g., AES-NI for AES, it is ideal for constructing other cryptographic primitives based on block ciphers.

1.1 Xor of Two Permutation Outputs

Bellare, Krovetz, and Rogaway [5] and Hall et al. [26] pioneered the investigation of constructing beyond-birthday-bound secure PRFs from PRPs, which has since attracted considerable attention [5, 26, 35, 36, 31, 19, 8, 24, 14, 25, 11]. One of the most well-known such constructions is so-called the xor of two permutations. Given a n-bit (keyed) PRP P, XoP maps $x \in \{0,1\}^{n-1}$ to

$$\mathsf{XoP}[\mathsf{P}](x) \stackrel{\text{def}}{=} \mathsf{P}(0 \parallel x) \oplus \mathsf{P}(1 \parallel x).$$

Alternatively, given two n-bit (keyed) PRPs P and Q, their sum, denoted $\mathsf{XoP2}$, maps $x \in \{0,1\}^n$ to

$$\mathsf{XoP2}[\mathsf{P}, \mathsf{Q}](x) \stackrel{\text{def}}{=} \mathsf{P}(x) \oplus \mathsf{Q}(x).$$

After the initial introduction of XoP construction [5, 26], several studies have built upon and enhanced this groundbreaking work [1, 18, 32, 34]. The most notable advancements include proofs by Dai, Hoang, and Tessaro [19] and Dutta, Nandi, and Saha [22], which established that XoP and $\mathsf{XoP2}$ are secure up to $O(2^n)$ queries. The two works use the chi-squared method and a verifiable version of mirror theory, respectively. However, their concrete security bounds are different; The tight bound of XoP is $\frac{q}{2^n}$ while the best known bound of $\mathsf{XoP2}$ is $O\left(\frac{q^2}{2^{2n}}\right)$ where q is the number of queries made by an adversary.

The difference can be more significant in the multi-user model. In the multi-user setting, Choi et al. [13] and Chen, Choi, and Lee [11] improved the multi-user security bound of $\mathsf{XoP2}$. Their result implies that if there are $O(2^n)$ number of XoP instances, i.e., $O(2^n)$ users, only one query per instance suffices to break PRF security of XoP. On the other hand, $\mathsf{XoP2}$ still enjoys beyond-birthday-bound in the same case.

1.2 MAC constructions

Nonce-Enhanced Hash-Then-MAC. On the other side, Dutta, Nandi, and Talnikar [23] presented an efficient construction of a message authentication code (MAC) called nonce-enhanced hash-then-MAC (nEHtM), achieving the BBB security both as a PRF and a MAC. Furthermore, this construction provides graceful security degradation of nonce misuse and only uses a (two-call of) single-block cipher and a single-block hash function such as the polynomial hash, making it a preferable option. The original construction of nEHtM is of the form:

$$\mathsf{nEHtM}[\mathsf{H},\mathsf{P}](N,M) \stackrel{\text{def}}{=} \mathsf{P}(0 \,\|\, N) \oplus \mathsf{P}(1 \,\|\, \mathsf{H}_{K_h}(M) \oplus N)$$

for a permutation P and appropriate hash function H.

The original paper proved the single-user security of nEHtM up to $O(2^{2n/3})$ MAC queries and $O(2^n)$ verification queries when the number of faulty queries is sufficiently small. Choi et al. [16] later improved this upto $O(2^{3n/4})$ MAC queries and $O(2^n)$ verification queries.

More recently, a variant of nEHtM has been considered, defined as

$$\mathsf{nEHtM2}[\mathsf{H},\mathsf{P},\mathsf{Q}](N,M) \stackrel{\text{def}}{=} \mathsf{P}(N) \oplus \mathsf{Q}(\mathsf{H}_{K_h}(M) \oplus N)$$

using two permutations, which we refer to $\mathsf{nEHtM2}$. This was first considered by Chen, Mennink, and Preneel [12], showing the single-user PRF security of this variant up to $O(2^{3n/4})$ queries. Chen, Choi, and Lee [11] proved that $\mathsf{nEHtM2}$ achieves stronger PRF security in the multi-user setting than the original nEHtM. In particular, they showed the BBB PRF security of $\mathsf{nEHtM2}$ for the number of users is about $2^{n/2}$, which was impossible for the original nEHtM because of the $uq/2^n$ term in the advantage bound. The (improved) MAC security of nEHtM and its variant in the multi-use setting is, on the other hand, left as an open problem.

Double-block Hash-then-Sum. Double-block Hash-then-Sum (DbHtS) paradigm was proposed by Datta et al. [20]. DbHtS has two versions: the two-key version based on $\mathsf{XoP1}$ and the three-key version based on $\mathsf{XoP2}$ where the number of keys implies the total number of keys used in a $2n$-bit output hash function and an n-bit block cipher. In this paper, we will focus on a variant of the two-key version of DbHtS, defined as follows: Let $\mathsf{H} = (\mathsf{H}^1, \mathsf{H}^2) : \{0,1\}^{2k} \times \mathcal{M} \to \{0,1\}^{n-1} \times \{0,1\}^{n-1}$ be a $(2n-2)$-bit hash function. H can be decomposed into two $(n-1)$-bit hash functions $\mathsf{H}^1, \mathsf{H}^2 : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^{n-1}$, and thus have $\mathsf{H}_{K_h}(M) = (\mathsf{H}^1_{K_{h_1}}(M), \mathsf{H}^2_{K_{h_2}}(M))$ where $K_h = (K_{h_1}, K_{h_2}) \in \{0,1\}^k \times \{0,1\}^k$. Let $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher modeled as an ideal cipher. We define (a variant of) the DbHtS constructions as follows

$$\mathsf{DbHtS}[\mathsf{H},\mathsf{E}](K_h,K,M) \stackrel{\text{def}}{=} \mathsf{E}_K(0\|\mathsf{H}^1_{K_{h,1}}(M)) \oplus \mathsf{E}_K(1\|\mathsf{H}^2_{K_{h,2}}(M)).$$

Note here we drop the least significant bit of the last round output of E deployed by $\mathsf{H}^1, \mathsf{H}^2$ so that both H^1 and H^2 outputs are $n-1$ bits, which is differentiated from previous works. We assume $\mathsf{H}^1, \mathsf{H}^2$ are δ_1-regular and δ_2-almost universal for $(n-1)$-bit outputs.

There are several works improved security analysis of DbHtS [29, 37, 21]; Notably, two-keyed DbHtS are proven to be tightly secure with a multi-user bound $O\left(\frac{\ell q^{4/3}}{2^n}\right)$ in the ideal cipher model when ℓ is the maximum length of messages, a $\frac{\ell}{2^n}$-universal (and regular) hash is used and discarding primitive queries by Datta et al. [21]. Those works on two-keyed DbHtS paradigm [37, 21] did not focus on domain separation and used the same block cipher call for each hash output; however, they required an additional hash property, namely cross-collision resistant, which can be realized by introducing domain separation.

1.3 Exaggerated Assumption in MAC Security Notion

Message authentication codes are widely accepted symmetric-key constructions. To ensure the security of MACs, we often compare those constructions with uniformly random functions. PRF security itself is not directly related to MAC security, but the indistinguishability from the uniform random functions is often used as an intermediate step for the security proof. In turn, the uniform random functions are somehow considered as *the ideal world*.

Albeit the ideal worlds are often identified as random functions or include random functions as a part of their interface, we notice that the random functions for MAC security do not need to be chosen uniformly randomly from the set of all possible functions: It does not rely on the all-zero output, denoted by $\mathbf{0}$, in the ideal worlds—they were introduced only to make analyses easier. Indeed, we can expect improvement by modifying the ideal worlds from the proof of [19]. While they introduced fine-tuned ideal worlds as intermediate worlds, their security bounds could not surpass $O\left(\frac{q}{2^n}\right)$ due to the presence of a trivial bad event that outputs $\mathbf{0}$ in the vanilla ideal world. This observation can lead to immediate improvement of security bounds of XoP-based constructions such as nEHtM [23, 16, 11] by the rid of the overestimated assumption. From this observation, we also newly introduce a variant of DbHtS [20, 29, 37, 21], which uses one block cipher key and domain separation.

1.4 Our Contribution

Our main contribution is, as stated above, observing the unnecessary loss in the previous security analyses and providing improved analyses in various models by introducing fine-tuning ideal worlds for those constructions: XoP, nEHtM, and (a variant of) DbHtS. We denote u as the number of users and q_m as the maximum number of queries per user allowed to an adversary. We use the standard model for XoP and nEHtM and the ideal cipher model for DbHtS to show that our observation can be applied regardless of the choice of models and the proof strategies. In the ideal cipher model, p stands for the number of primitive queries allowed to an adversary.

Security of XoP. We show that the "fine-tuned" multi-user PRF security bound of XoP from the random ideal world without outputting zero can be

1. $O\left(\frac{u^{0.5}{q_m}^{1.5}}{2^{1.5n}}\right)$ via the Chi-squared method [19] where the same security bound for $\mathsf{XoP2}$ was proven in Choi et al. [13] at ASIACRYPT'22;
2. $O\left(\frac{u^{0.5}{q_m}^{2}}{2^{2n}}\right)$ via the Squared-ratio method [11] where the same security bound for $\mathsf{XoP2}$ was proven in the same paper at CRYPTO'23.

Note that just checking if there is an output $\mathbf{0}$ of the oracle breaks the standard PRF security for $q \geq 2^n$, making no hope for better than $O\left(\frac{uq_m}{2^n}\right)$ security. Our result for XoP demonstrates this barrier is entirely due to the output $\mathbf{0}$.

Security of nEHtM. We revisit the multi-user security of the original nEHtM in the multi-use setting. We prove that nEHtM enjoys strong MAC multi-user security similar to the multi-user PRF result in [11] while using less key size with graceful security degradation under nonce misuse, resolving the open question posed in [11]. When the number of users $u = O(2^{n/2})$ and each user makes the faulty queries much less than $2^{n/4}$ times, then our result indicates that nEHtM is BBB secure MAC. This was believed to be impossible, at least through the standard ideal world—with outputting zero. Concretely, we prove that the multi-user MAC security bound of nEHtM is $O\left(\left(\frac{uq_m^4}{2^{3n}}\right)^{1/2}\right)$ as long as the number of faulty and verification queries is sufficiently small and q_m is large enough. The previous best bound in a similar setting was $O\left(\frac{uq_m^2}{2^{1.5n}}\right)$ [16]. A similar security of nEHtM as PRF without outputting zero is also proven.

Along the way, we figure out that the multi-user PRF security of $\mathsf{nEHtM2}$ in [11] is buggy (see Sections 1.5 and 6.5), resulting in a slightly worse bound than they claimed; for example, the claimed birthday bound security for $u \approx 2^n$ is false. Despite this, we develop and fine-tune the relevant extended mirror theory without outputting zero and the security proof of nEHtM, resulting in the *even better bound* than one for $\mathsf{nEHtM2}$ in [11] in some sense. For example, their security bound does not work for $q_m \approx 2^{3n/4}$. We refer to Figure 2 for the graphical comparison.

We also study variants of nEHtM and $\mathsf{nEHtM2}$ based on a stronger hash function. This variant (almost) recovers the security multi-used PRF claim for $\mathsf{nEHtM2}$ in [11] if we use their original proof, and the even better MAC security bound of nEHtM including $O\left(\frac{u^{0.5}q_m^4}{2^{3n}}\right)$ if we exploit the improved strategies and mirror theory in this paper. This exhibits the power of our fine-tuning and indicates that the current obstacles to better and cleaner security are from the hash functions, either its property itself or its current analysis.

Security of DbHtS. At last, we explore the multi-user MAC security of DbHtS. Our main targets are the variants of [21, 37] using the domain separation. We focus on the security bound that is fine-tuned in terms of the query bound q_m for each user, instead of the total number of queries q across all users. In the worst case, $q = uq_m$ holds. Our results can be summarized as follows.

– Under the ideal cipher model as in [37], we analyze the security of DbHtS based on a dedicated analysis regarding q_m along with the idea of fine-tuning but mainly following the original approach. This leads to a better security bound than one by the so-called generic reduction and also achieves an improvement over the original result in the same setting, except for the domain separation.
– For [21], if we focus on q_m, we observe that naïvely following the original proof cannot avoid $uq_m^{4/3}/2^n$, which is even worse than the trivial bad probability of $uq_m/2^n$. Inspired by the case of nEHtM, we show that an improved bound can be achieved assuming stronger underlying hash functions.

A pictorial comparison is shown in Figure 1. The effect of fine-tuning also appears in the low end, for example, when $q_m \lesssim 2^{n/3}$ still allows the $1/2^n$ security bound in both cases when the other parameters are sufficiently small, which was impossible in the previous bounds.

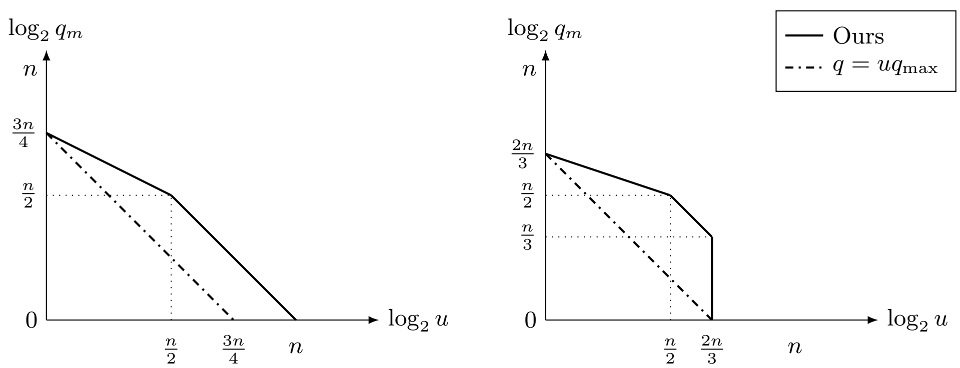

Fig. 1: Comparison of the security bounds (in terms of the threshold number of queries per user) as functions of $\log_2 u$. The solid line represents our bounds, and the dash-dotted line represents the previous bound where $q = uq_{\max}$. The left figure compares our Theorem 13 with Theorem 1 from [21]. The right figure compares our Theorem 12 with Theorem 1 from [37], where we set $\epsilon_3, \epsilon_4 = O(2^{-2n})$ according to [37, p. 19]. We set $p = 0, k = n, \delta = 2^{-n}$, and neglect l and the logarithmic term of n in all graphs.

We remark that the use of strong hash functions for better security is reminiscent of the recent advances in the cascaded LRW2 security. Mennink [33] presented an attack on Cascaded LRW2 [30], or CLRW2, and showed that matching security bound, *under several assumptions*, including a stronger property of hash functions about the multiple collisions, very similar to ours. Subsequently, Jha and Nandi [28] demonstrated how to eliminate those assumptions and developed a couple of new tools for dealing with the multiple collisions of the hash functions, apparently inspired by the result of Mennink. In turn, these tools are frequently used in the later works [11, 16, 12] as well as this work. We hope our security bounds with strong hashes highlight the specific point in the security proofs that future works resolve.

1.5 The security bound of nEHtM2 in [11]

We briefly sketch the problems in the original multi-user nEHtM2 security proof in [11], confirmed by private communications with the authors. We stress that the main problems are from the security proof of nEHtM itself, not from their main tool, the Squared-ratio method and Mirror theory.

The main issue is the behavior of the property of hash functions H and q_c, the number of edges in the components of size > 2 in the graph representation of queries. In the nonce-respecting setting, q_c increases when $X_i = X_j$ happens for $X_i = \mathsf{H}_{K_h}(M_i) \oplus N_i$ for the message M_i and nonce N_i. Since H is δ-almost XOR universal, we can only predict the property of single event $X_i = X_j$ that is paraphrased by $\mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) = N_i \oplus N_j$. This suffices for estimating the expectation of q_c. However, we need to compute the expectation of q_c^2 and give the bound on q_c with a high probability. Computing $\mathbf{Ex}\left[q_c^2\right]$ is involved with *multiple* collisions, i.e., the event that $X_i = X_j$ and $X_k = X_\ell$ simultaneously happen. [11] implicitly assumes that two collisions happen independently, giving a nice upper bound of $\mathbf{Ex}\left[q_c^2\right]$ (as in Fact 6 derived using stronger hash functions). However, what we can actually give is a worse bound as in Fact 5. They again use their false estimation when computing the probability for some bad event (bad_5 in their proof).

Another minor issue is a missing term at the end of the proof. In $\mathsf{Adv}^{\mathsf{mu-prf}}_{\mathsf{nEHtM}}$ bound in [11, page 28], there is a term about $\frac{\sqrt{u}nLq_{\max}\delta}{2^n}$ at the second line. However, there is no corresponding term in the final bound. A similar term $\frac{\sqrt{u}nLq_{\max}^2\delta}{2^n}$ is alive, which is smaller than the problematic term when $q_{\max}^2\delta \leq 1$, This missing part affects some exponent of the final security bound.

We show that using a stronger hash function allows us to recover a similar result as the original. We refer to Section 6.5 for a more detailed analysis.

1.6 Version Notes

Some parts of this paper have been revised from the original version to improve the presentations and some technical parts. We summarize the differences below.

The Eurocrypt submission version. This is the original version.
The Eprint (2023-Oct.) version.
- We make the consistency between the presentations in each section, and correct and improve many presentations.
- We add Lemmas 4 and 6 as separated lemmas.
- Proofs of Mirror theory are revised.
- We apply Lemma 6 in proving Theorem 10 for a cleaner presentation, though it has a slightly worse constant.

– In Section 6, we give a colorful transition of equations for easier verification. We also made some changes at 1) the event bad_6 by $q_c \geq \frac{2^{2n}}{186q_m^2}$ and the conditions of L_1, L_2 (originally it was $q_c \geq \min\left(\frac{2^{2n}}{186q_m^2}, \frac{2^n}{3(5L_1+5L_2+2\mu_m)}\right)$), 2) the analysis of bad_{3c} based on $\neg\mathsf{bad}_{2a}$ (originally it follows the analysis of [16, bad_{2e}]), and 3) the choice of L_2, which now considers two cases. These changes improve the final bound, allowing the security makes sense for $u \lesssim 2^{2n}$, while the previous bound only works for $u \lesssim 2^{\frac{5n}{3}}$. We also slightly improve the computations of $\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]$ at the beginning of Section 6.2, and fill the sanity check for $\xi_{\max} q_m \ll 2^n$ which was not explicitly written in the original version.
– We modify Figure 2; we adjust the improvement in this version for our result, and correct an error of the graph for [11] in the previous version. We also add the graphs when assuming δ-AXU$^{(2)}$.
– We provide a rigorous proof of Theorem 13. In the proof of Theorem 13, we replace the misused mirror theory lower bound ([21, Lemma 1]) with our fine-tuning extended mirror theory lower bound (Theorem 6). The correction doesn't affect the dominant term in the statement.

2 Preliminaries

Notation. Throughout this paper, we fix positive integers n and u to denote the block size and the number of users, respectively. For a non-empty finite set $\mathcal{X}$, we let $\mathcal{X}^{*\ell}$ denote a set $\{(x_1, \ldots, x_\ell) \in \mathcal{X}^\ell \mid x_i \neq x_j \text{ for } i \neq j\}$. For an integer A and b, we denote $(A)_b = A(A-1)\ldots(A-b+1)$. A notation $x \leftarrow_\$ \mathcal{X}$ means that x is chosen uniformly at random from $\mathcal{X}$. $|\mathcal{X}|$ means the number of elements in $\mathcal{X}$. The set of all permutations of $\{0,1\}^n$ is simply denoted $\mathsf{Perm}(n)$. The set of all functions with domain $\{0,1\}^n$ and codomain $\{0,1\}^m$ is simply denoted by $\mathsf{Func}(n,m)$. We additionally define $\mathsf{Func}^*(n,m) \subset \mathsf{Func}(n,m)$ by the set of all functions in $\mathsf{Func}(n,m)$ satisfying the following condition: for any $f \in \mathsf{Func}^*(n,m)$, $f(x) \neq \mathbf{0}$ for all $x \in \{0,1\}^n$. For a keyed function $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ with key space $\mathcal{K}$, and non-empty sets $\mathcal{X}$ and $\mathcal{Y}$, we will denote $F(K, \cdot)$ by $F_K(\cdot)$ for $K \in \mathcal{K}$. When two sets $\mathcal{X}$ and $\mathcal{Y}$ are disjoint, their (disjoint) union is denoted $\mathcal{X} \sqcup \mathcal{Y}$. We write T_{re} and T_{id} as random variables following the distribution of the transcripts in the real world and the ideal world, respectively. For any positive integer i, and $a_1, \ldots, a_i, b \in \{0,1\}^n$, We denote $\{a_1, \ldots, a_i\} \oplus b \stackrel{\text{def}}{=} \{a_1 \oplus b, \ldots, a_i \oplus b\}$

Almost XOR Universal Hash Functions. Let $\delta > 0$, and let $\mathsf{H} : \mathcal{K}_h \times \mathcal{M} \to \mathcal{X}$ be a keyed function for three non-empty sets $\mathcal{K}_h$, $\mathcal{M}$, and $\mathcal{X}$. We say that H is *δ-XOR almost universal* (δ-XAU) if for any distinct $M, M' \in \mathcal{M}$ and $X \in \mathcal{X}$,

$$\Pr\left[K_h \leftarrow_\$ \mathcal{K}_h : \mathsf{H}_{K_h}(M) \oplus \mathsf{H}_{K_h}(M') = X\right] \leq \delta.$$

Regular and Almost Universal Hash Functions. Let $\delta_1, \delta_2 > 0$, and let $\mathsf{H} : \mathcal{K}_h \times \mathcal{M} \to \mathcal{X}$ be a keyed function for three non-empty sets $\mathcal{K}_h$, $\mathcal{M}$, and $\mathcal{X}$.

We say that H is δ_1-regular if for any $M \in \mathcal{M}$ and $X \in \mathcal{X}$,

$$\Pr\left[K_h \leftarrow_{\$} \mathcal{K}_h : \mathsf{H}_{K_h}(M) = X\right] \leq \delta_1,$$

and H is δ_2 *almost universal* (δ_2-AU) if for any distinct $M, M' \in \mathcal{M}$ and $X \in \mathcal{X}$,

$$\Pr\left[K_h \leftarrow_{\$} \mathcal{K}_h : \mathsf{H}_{K_h}(M) = \mathsf{H}_{K_h}(M')\right] \leq \delta_2.$$

2.1 The Chi-Squared Method

We give here the necessary background on the chi-squared method [19].

We fix a set of random systems, a deterministic distinguisher $\mathcal{A}$ that makes q oracle queries to one of the random systems, and a set Ω that contains all possible answers for oracle queries to the random systems. For a random system $\mathcal{S}$ and $i \in \{1, \ldots, q\}$, let $Z_{\mathcal{S},i}$ be the random variable over Ω that follows the distribution of the i-th answer obtained by $\mathcal{A}$ interacting with $\mathcal{S}$. Let

$$\mathbf{Z}_{\mathcal{S}}^{i} \stackrel{\text{def}}{=} (Z_{\mathcal{S},1}, \ldots, Z_{\mathcal{S},i}),$$

and let

$$\mathsf{p}_{\mathcal{S}}^{i}(\mathbf{z}) \stackrel{\text{def}}{=} \Pr\left[\mathbf{Z}_{\mathcal{S}}^{i} = \mathbf{z}\right]$$

for $\mathbf{z} \in \Omega^i$. We omit i when $i = q$. For $i \leq q$ and $\mathbf{z} = (z_1, \ldots, z_{i-1}) \in \Omega^{i-1}$ such that $\mathsf{p}_{\mathcal{S}}^{i-1}(\mathbf{z}) > 0$, the probability distribution of $Z_{\mathcal{S},i}$ conditioned on $\mathbf{Z}_{\mathcal{S}}^{i-1} = \mathbf{z}$ will be denoted $\mathsf{p}_{\mathcal{S},i}^{\mathbf{z}}(\cdot)$, namely for $z \in \Omega$,

$$\mathsf{p}_{\mathcal{S},i}^{\mathbf{z}}(z) \stackrel{\text{def}}{=} \Pr\left[Z_{\mathcal{S},i} = z \mid \mathbf{Z}_{\mathcal{S}}^{i-1} = \mathbf{z}\right].$$

For two random systems $\mathcal{S}_0$ and $\mathcal{S}_1$, and for $i < q$ and $\mathbf{z} = (z_1, \ldots, z_{i-1}) \in \Omega^{i-1}$ such that $\mathsf{p}_{\mathcal{S}_0}^{i-1}(\mathbf{z})$, $\mathsf{p}_{\mathcal{S}_1}^{i-1}(\mathbf{z}) > 0$, the χ^2-*divergence* for $\mathsf{p}_{\mathcal{S}_0,i}^{\mathbf{z}}(\cdot)$ and $\mathsf{p}_{\mathcal{S}_1,i}^{\mathbf{z}}(\cdot)$ is defined as follows.

$$\chi^2\left(\mathsf{p}_{\mathcal{S}_1,i}^{\mathbf{z}}(\cdot), \mathsf{p}_{\mathcal{S}_0,i}^{\mathbf{z}}(\cdot)\right) \stackrel{\text{def}}{=} \sum_{\substack{z \in \Omega \text{ such that} \\ \mathsf{p}_{\mathcal{S}_0,i}^{\mathbf{z}}(z) > 0}} \frac{\left(\mathsf{p}_{\mathcal{S}_1,i}^{\mathbf{z}}(z) - \mathsf{p}_{\mathcal{S}_0,i}^{\mathbf{z}}(z)\right)^2}{\mathsf{p}_{\mathcal{S}_0,i}^{\mathbf{z}}(z)}.$$

We will simply write $\chi^2(\mathbf{z}) = \chi^2\left(\mathsf{p}_{\mathcal{S}_1,i}^{\mathbf{z}}(\cdot), \mathsf{p}_{\mathcal{S}_0,i}^{\mathbf{z}}(\cdot)\right)$ when the random systems are clear from the context. If the support of $\mathsf{p}_{\mathcal{S}_1}^{i-1}(\cdot)$ is contained in the support of $\mathsf{p}_{\mathcal{S}_0}^{i-1}(\cdot)$, then we can view $\chi^2\left(\mathsf{p}_{\mathcal{S}_1,i}^{\mathbf{z}}(\cdot), \mathsf{p}_{\mathcal{S}_0,i}^{\mathbf{z}}(\cdot)\right)$ as a random variable, denoted $\chi^2\left(\mathbf{Z}_{\mathcal{S}_1}^{i-1}\right)$, where $\mathbf{z}$ follows the distribution of $\mathbf{Z}_{\mathcal{S}_1}^{i-1}$.

Then $\mathcal{A}$'s distinguishing advantage is upper bounded by the *total variation distance* of $\mathsf{p}_{\mathcal{S}_0}(\cdot)$ and $\mathsf{p}_{\mathcal{S}_1}(\cdot)$, denoted $\|\mathsf{p}_{\mathcal{S}_0}(\cdot) - \mathsf{p}_{\mathcal{S}_1}(\cdot)\|$, and we have the following theorem.

Theorem 1 ([19]). *Suppose whenever* $\mathsf{p}_{\mathcal{S}_1^1}(\cdot) > 0$ *then* $\mathsf{p}_{\mathcal{S}_0^1}(\cdot) > 0$. *Then we have*

$$\|\mathsf{p}_{\mathcal{S}_0}(\cdot) - \mathsf{p}_{\mathcal{S}_1}(\cdot)\| \leq \left(\frac{1}{2}\sum_{i=1}^{q} \mathbf{Ex}\left[\chi^2\left(\mathbf{Z}_{\mathcal{S}_1}^{i-1}\right)\right]\right)^{\frac{1}{2}}. \tag{1}$$

2.2 The Squared-Ratio Method

This method was first introduced in [11]. For multi-user security, we assume a random system $\mathcal{S} = (\mathcal{S}^1, \ldots, \mathcal{S}^u)$ and $i \in \{1, \ldots, u\}$. Note that $\mathcal{S}_0^1$ is the ideal world and $\mathcal{S}_1^1$ is the real world for the first user, independent of the other user's oracle.

Theorem 2 ([11]). *Suppose whenever $\mathsf{p}_{\mathcal{S}_1^1}(\cdot) > 0$ then $\mathsf{p}_{\mathcal{S}_0^1}(\cdot) > 0$. Let $\Omega = \Gamma_{\mathsf{good}} \sqcup \Gamma_{\mathsf{bad}}$. If a function $\epsilon_1(\mathbf{z})$ and a constant ϵ_2 holds the following constraints*

$$\left| \frac{\mathsf{p}_{\mathcal{S}_1^1}(\mathbf{z})}{\mathsf{p}_{\mathcal{S}_0^1}(\mathbf{z})} - 1 \right| \leq \epsilon_1(\mathbf{z})$$

for all attainable $\mathbf{z} \in \Gamma_{\mathsf{good}}$ and

$$\Pr\left[Z_{\mathcal{S}_0^1} \in \Gamma_{\mathsf{bad}} \right] \leq \epsilon_2,$$

one has

$$\|\mathsf{p}_{\mathcal{S}_1}(\cdot) - \mathsf{p}_{\mathcal{S}_0}(\cdot)\| \leq \sqrt{2u\mathbf{Ex}\left[\epsilon_1(\mathbf{z})^2\right]} + 2u\epsilon_2$$

where the expectation is taken over the distribution of $Z_{\mathcal{S}_0^1}$.

Note that the expectation in the squared-ratio method is the distribution from the ideal world, while the expectation in the Chi-squared method is taken over the distribution from the real world.

2.3 Patarin's H-Coefficient Technique

The well-known Patarin's H-coefficient technique can be expressed as below:

Lemma 1 ([10]). *Suppose whenever $\mathsf{p}_{\mathcal{S}_1^1}(\cdot) > 0$ then $\mathsf{p}_{\mathcal{S}_0^1}(\cdot) > 0$. Let $\Omega = \Gamma_{\mathsf{good}} \sqcup \Gamma_{\mathsf{bad}}$. Let $\epsilon_1, \epsilon_2 \geq 0$ be two constants. If $\frac{\mathsf{p}_{\mathcal{S}_1^1}(z)}{\mathsf{p}_{\mathcal{S}_0^1}(z)} \geq 1 - \epsilon_1$ holds for all attainable $z \in \Gamma_{\mathsf{good}}$ and $\Pr\left[Z_{\mathcal{S}_0^1} \in \Gamma_{\mathsf{bad}} \right] \leq \epsilon_2$. Then, it holds that*

$$\|\mathsf{p}_{\mathcal{S}_1^1}(\cdot) - \mathsf{p}_{\mathcal{S}_0^1}(\cdot)\| \leq \epsilon_1 + \epsilon_2,$$

where the expectation is taken over the distribution of $Z_{\mathcal{S}_0^1}$.

2.4 Useful Inequalities

We use the following inequalities multiple times in the proof.

$$\prod_{i=1}^{n}(1 - x_i) \geq 1 - \sum_{i=1}^{n} x_i \text{ if } 0 \leq x_i \leq 1 \text{ for all } i \tag{2}$$

$$\sum_{i=1}^{n} i^k \geq \frac{n^{k+1}}{k+1} \text{ for } k \geq 1 \tag{3}$$

Lemma 2 **(Markov's inequality).** *Let X be a non-negative random variable and $a > 0$. It holds that*

$$\Pr[X \geq a] \leq \mathbf{Ex}[X]/a.$$

Lemma 3 **(Chebyshev's inequality).** *Let X be a random variable and $t > 0$. It holds that*

$$\Pr[X \geq \mathbf{Ex}[X] + t] \leq \frac{\mathbf{Var}[X]}{t^2}.$$

Lemma 4 **(Bonferroni's inequality).** *For events $A_1, ..., A_n$, it holds that*

$$\sum_{i=1}^{n} \Pr[A_i] - \sum_{1 \leq i < j \leq n} \Pr[A_i \wedge A_j] \leq \Pr[\vee_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i].$$

The upper bound is usually called the union bound.

Lemma 5. *Let M, q be positive integers. If $(\lambda_1, ..., \lambda_q) \in [M]^q$ are uniformly randomly distributed, then the number of collisions*

$$C = \left|\{(i,j) \in [q]^2 : (i < j) \wedge (\lambda_i = \lambda_j)\}\right|$$

satisfies the following inequalities hold for any $t > 0$:

$$\mathbf{Ex}[C] \leq \frac{q^2}{2M}, \quad \mathbf{Var}[C] \leq \frac{q^2}{2M}, \quad \Pr\left[C \geq \frac{q^2}{2M} + t\right] \leq \frac{q^2}{2Mt^2}.$$

Furthermore, if $q^2 < 2M$, it also holds that $\mathbf{Ex}\left[C^2\right] \leq q^2/M$, and $\mathbf{Ex}\left[C^2\right] \leq q^4/2M^2$ otherwise.

Proof. Let $I_{i,j}$ equal 1 if $\lambda_i = \lambda_j$, and 0 otherwise. It holds that $\mathbf{Ex}[I_{i,j}] = 1/M$ and $C = \sum_{i<j} I_{i,j}$, and

$$\mathbf{Ex}[C] = \sum_{i<j\leq q} \mathbf{Ex}[I_{i,j}] = \frac{q(q-1)}{2M} \leq \frac{q^2}{2M}.$$

For the variance, it holds that

$$\begin{aligned}\mathbf{Var}[C] = \mathbf{Var}[\sum_{i<j} I_{i,j}] &= \mathbf{Ex}\left[\left(\sum_{i<j}\left(I_{i,j} - \frac{1}{M}\right)\right)^2\right] \\ &= \sum_{i<j}\sum_{k<\ell} \mathbf{Ex}\left[\left(I_{i,j} - \frac{1}{M}\right)\left(I_{k,\ell} - \frac{1}{M}\right)\right] = \sum_{i<j}\sum_{k<\ell} \mathbf{Ex}\left[I_{i,j}I_{k,\ell} - \frac{1}{M^2}\right]\end{aligned}$$

We consider two cases as follows: 1) $(i,j) = (k,\ell)$, then $\mathbf{Ex}[I_{i,j}I_{k,\ell}] = 1/M$ with $\binom{q}{2}$ possible choices, and 2) $|\{i,j\} \cap \{k,\ell\}| \leq 1$, then $\mathbf{Ex}[I_{i,j}I_{k,\ell}] = 1/M^2$ anyway and the relevant term becomes zero. Overall, it holds that

$$\mathbf{Var}[C] = \binom{q}{2}\left(\frac{1}{M} - \frac{1}{M^2}\right) \leq \frac{q^2}{2M},$$

and finally $\mathbf{Ex}\left[C^2\right] = \mathbf{Var}[C] + \mathbf{Ex}[C]^2$ gives the final statement. $\square$

Lemma 6. *For $X_1, ..., X_k$, it holds that*

$$\sqrt{\left(\sum_{i=1}^{k} X_i\right)^2} \le \sqrt{k \cdot \sum_{i=1}^{k} X_i^2} \le \sum_{i=1}^{k} \sqrt{k \cdot X_i{}^2}.$$

In particular, it holds that

$$\sqrt{\mathbf{Ex}\left[\left(\sum_{i=1}^{k} X_i\right)^2\right]} \le \sqrt{k\mathbf{Ex}\left[\sum_{i=1}^{k} X_i^2\right]} \le \sum_{i=1}^{k} \sqrt{k\mathbf{Ex}\left[X_i^2\right]},$$

The first inequality is due to Cauchy-Schwartz inequality, and the second is obvious. Although this is usually not tight (but only loss a constant factor for constant k), we use it in the squared-ratio method for deriving a simple upper bound of $\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]$.

3 Fine-Tuning Security Notions

We define a new fine-tuned multi-user pseudorandom function security from the function domain $\mathsf{Func}^*(m, n)$ instead of $\mathsf{Func}(m, n)$.

Pseudorandom Functions without 0^n. Let $\mathsf{C} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^m$ be a keyed function with key space $\mathcal{K}$. We will consider an (information-theoretic) distinguisher $\mathcal{A}$ that makes oracle queries to C_{K_i} for multiple keys K_i for $i \in [u]$ and returns a single bit. The advantage of $\mathcal{A}$ in breaking the $\mathsf{mu\text{-}prf}^*$ security of C, i.e., in distinguishing $\mathsf{C}(K_1, \cdot), \ldots, \mathsf{C}(K_u, \cdot)$ where $K_1, \ldots, K_u \leftarrow_\$ \mathcal{K}$ from uniformly chosen functions $\mathsf{F}_1, \ldots, \mathsf{F}_u \leftarrow_\$ \mathsf{Func}^*(n, m)$, is defined as

$$\begin{aligned}\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(\mathcal{A}) = \Big|&\Pr\left[K_1, \ldots, K_u \leftarrow_\$ \mathcal{K} : \mathcal{A}^{\mathsf{C}_{K_1}(\cdot), \ldots, \mathsf{C}_{K_u}(\cdot)} = 1\right] \\ &- \Pr\left[\mathsf{F}_1, \ldots, \mathsf{F}_u \leftarrow_\$ \mathsf{Func}^*(n, m) : \mathcal{A}^{\mathsf{F}_1(\cdot), \ldots, \mathsf{F}_u(\cdot)} = 1\right]\Big|.\end{aligned}$$

We define $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(u, q_m, t)$ as the maximum of $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(\mathcal{A})$ over all the distinguishers against C for u users making at most q_m queries to each user and running in time at most t. When we consider information-theoretic security, we will drop the parameter t.

When the global primitive—usually the ideal cipher—is given to an adversary, we take into account the number of queries to the primitive oracle made by the adversary. We define $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(u, q_m, p)$ as the maximum of $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(\mathcal{D})$ over all the distinguishers against C making at most q_m construction queries to each of u users and p primitive queries in total.

Nonce-based MACs. Given four non-empty sets $\mathcal{K}$, $\mathcal{N}$, $\mathcal{M}$, and $\mathcal{T}$, a nonce-based keyed function with key space $\mathcal{K}$, nonce space $\mathcal{N}$, message space $\mathcal{M}$ and tag space $\mathcal{T}$ is a function $F : \mathcal{K} \times \mathcal{N} \times \mathcal{M} \to \mathcal{T}$. Stated otherwise, it is a keyed function whose domain is a cartesian product $\mathcal{N} \times \mathcal{M}$. We will sometimes write $F_K(N, M)$ to denote $F(K, N, M)$.

For $K \in \mathcal{K}$, let Auth_K be the MAC oracle which takes as input a pair $(N, M) \in \mathcal{N} \times \mathcal{M}$ and returns $F_K(N, M)$, and let Ver_K be the verification oracle which takes as input a triple $(N, M, T) \in \mathcal{N} \times \mathcal{M} \times \mathcal{T}$ and returns $\top$ ("accept") if $F_K(N, M) = T$, and $\bot$ ("reject") otherwise. We assume that an adversary makes queries to the two oracles Auth_K and Ver_K for a secret key $K \in \mathcal{K}$. Assuming that, without loss of generality, an adversary never makes the verification query that it received from the MAC query, we say that an adversary forges if its queries to the oracle Ver_K returns $\top$ for some K. A MAC query (N, M) made by an adversary is called a *faulty query* if the adversary has already queried the MAC oracle with the same nonce but with a different message; we sometimes call both the faulty query and the corresponding previous query with the same nonce by a query with a repeated nonce.

In the multi-user setting, a (u, μ_m, q_m, v_m, t)-adversary against the nonce-based MAC security of F is an algorithm $\mathcal{A}$ with oracle access to Auth_{K_i} and Ver_{K_i} for $i \in [i]$, making at most q_m MAC queries, at most μ_m faulty queries, and at most v_m verification queries to Auth_{K_i} and Ver_{K_i} for each $i \in [i]$, and runs in time at most t. The multi-user MAC advantage of $\mathcal{F}$ against (u, μ_m, q_m, v_m, t)-adversary, denoted by $\mathsf{Adv}_{\mathsf{F}}^{\mathsf{mu\text{-}mac}}(u, \mu_m, q_m, v_m, t)$, is defined by

$$\max_{\mathcal{A}} \Pr\left[K_1, ..., K_u \leftarrow_{\$} \mathcal{K} : \mathcal{A}^{\mathsf{Auth}_{K_1}, \ldots, \mathsf{Auth}_{K_u}, \mathsf{Ver}_{K_1}, \ldots, \mathsf{Ver}_{K_u}} \text{ forges}\right]$$

where max is taken over all (u, μ_m, q_m, v_m, t)-adversary $\mathcal{A}$ against the nonce-based MAC security of F. We occasionally drop the parameter t when we focus on information-theoretic security. When $\mu_m = 0$, we say that $\mathcal{A}$ is nonce-respecting.

In this work, we prove the MAC security of F by comparing it with the ideal world of MAC. That is, we consider the ideal world oracles Rand_i^* and Rej_i for $i \in [u]$, where Rand_i^* returns an independent random value except 0^n, instantiated by a truly random function from $\mathsf{Func}^*(m, n)$, and Rej_i always returns $\bot$ for every verification query. Then, $\mathsf{Adv}_{\mathsf{F}}^{\mathsf{mu\text{-}mac}}(u, \mu_m, q_m, v_m, t)$ is bounded above by

$$\max_{\mathcal{D}} \Big| \Pr\left[K_1, ..., K_u \leftarrow_{\$} \mathcal{K} : \mathcal{D}^{\mathsf{Auth}_{K_1}, \ldots, \mathsf{Auth}_{K_u}, \mathsf{Ver}_{K_1}, \ldots, \mathsf{Ver}_{K_u}} = 1\right]$$
$$- \Pr\left[\mathcal{D}^{\mathsf{Rand}_1^*, \ldots, \mathsf{Rand}_u^*, \mathsf{Rej}_1, \ldots, \mathsf{Rej}_u} = 1\right] \Big|$$

where max is taken over all (u, μ_m, q_m, v_m, t)-adversary. This is easily proven by using forgery adversary $\mathcal{A}$ to distinguish the two worlds. In turn, we mainly focus on showing the above indistinguishability for MAC security.

4 Fine-Tuning Extended Mirror Theory with Upper Bounds

Definitions and Notations. We write $N = 2^n$ for simplicity. Let r, q, p be fixed nonnegative integers such that $r \leq 2(p + q)$. The sets $\mathcal{P} = \{P_1, ..., P_r\}$

of *unknown* variables $P_i \in \{0,1\}^n$ for $i \in [r]$, where $P_i \neq P_{i'}$ for $i \neq i'$. We consider two types of relations between variables, *equations* and *non-equations*. The system of equations is represented by a sequence of constants $(\lambda_1, ..., \lambda_q) \in (\{0,1\}^n)^q$ along with indices $\gamma_1, \ldots, \gamma_q, \gamma'_1, \ldots, \gamma'_q \in [r]$ such that $\gamma_i \neq \gamma'_{i'}$ for any $i, i' \in [r]$ and the equations

$$\Gamma^= : \begin{cases} P_{\gamma_1} \oplus P_{\gamma'_1} = \lambda_1, \\ P_{\gamma_3} \oplus P_{\gamma'_2} = \lambda_2, \\ \qquad \vdots \\ P_{\gamma_q} \oplus P_{\gamma'_q} = \lambda_q \end{cases}$$

hold. Similarly, a sequence of constants $(\mu_1, ..., \mu_p) \in (\{0,1\}^n)^p$ and indices $\sigma_1, \ldots, \sigma_p, \sigma'_1, \ldots, \sigma'_p \in [r]$ determine the system of inequations

$$\Gamma^{\neq} : \begin{cases} P_{\sigma_1} \oplus P_{\sigma'_1} \neq \mu_1, \\ P_{\sigma_2} \oplus P_{\sigma'_2} \neq \mu_2, \\ \qquad \vdots \\ P_{\sigma_p} \oplus P_{\sigma'_p} \neq \mu_p \end{cases}$$

where $\sigma_i \neq \sigma'_{i'}$ for any $i, i' \in [r]$ The overall system is denoted by Γ. When the variables in $\mathcal{P}$ are assigned by some values, we will identify the variables with the values assigned to them.

Graph-theoretic interpretation. Two systems $\Gamma = (\Gamma^=, \Gamma^{\neq})$ give corresponding simple graphs $\mathcal{G}^= = \mathcal{G}(\Gamma^=) = (\mathcal{P}, \mathcal{E}^=)$ and $\mathcal{G}^{\neq} = \mathcal{G}(\Gamma^{\neq}) = (\mathcal{P}, \mathcal{E}^{\neq})$. The sets of edges are defined by

$$\mathcal{E}^= = \{(P_{\gamma_i}, P_{\gamma'_i}) : i \in [q]\}, \quad \mathcal{E}^{\neq} = \{(P_{\sigma_i}, P_{\sigma'_i}) : i \in [p]\}.$$

Each edge $(P, P') \in \mathcal{E}^=$ is labeled by $(=, \lambda)$ if $P \oplus P' = \lambda$ is included in $\Gamma^=$ and $(P, P') \in \mathcal{E}^{\neq}$ is labeled by $(\neq, \mu)$ if $P \oplus P' \neq \mu$. We sometimes write $P \overset{\star}{-} P'$ when an edge (P, P') is labeled with $(=, \star)$, and define the label function λ by $\lambda(P, P') = \star$. We also define the function μ by $\mu(P, P') = \star$ if (P, P') is labeled with $(\neq, \star)$. Throughout this paper, we only consider the graph $\mathcal{G}^=$ with no loops, i.e., that is *acyclic*.

For the graph of equations $\mathcal{G}^=$, let $\mathcal{L}$ be a trail of ℓ-length

$$\mathcal{L} : V_0 \overset{\lambda_1}{-} V_1 \overset{\lambda_2}{-} \cdots \overset{\lambda_\ell}{-} V_\ell.$$

We can naturally extend λ to the trails by defining

$$\lambda(\mathcal{L}) \overset{\text{def}}{=} \lambda_1 \oplus \lambda_2 \oplus \cdots \oplus \lambda_\ell,$$

and we say that $\mathcal{L}$ is $\lambda(\mathcal{L})$-labeled. Since $\mathcal{G}^=$ is acyclic, $\lambda(V_0, V_\ell) \overset{\text{def}}{=} \lambda(\mathcal{L})$ is well-defined. If V and V' are not connected, we define $\lambda(V, V') = \perp$.

Recall that the equation graph $\mathcal{G}^=$ is acyclic. Also, since the variables in $\mathcal{P}$ take the different values, $\mathcal{G}^=$ must satisfy that $\lambda(\mathcal{L}) \neq \mathbf{0}$ for any trail $\mathcal{L}$ we say that the graph is *non-degenerated* if it satisfies this property. The union graph

$$\mathcal{G} = \mathcal{G}(\Gamma) = (\mathcal{V}, \mathcal{E}^= \cup \mathcal{E}^{\neq})$$

does not contain isolated vertices, i.e., every vertex has a positive degree.

We decompose the set of vertices $\mathcal{V}$ of the graph $\mathcal{G}^=$ into its connected components

$$\mathcal{V} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup ... \sqcup \mathcal{C}_{\alpha+\beta} \sqcup \mathcal{D} \tag{4}$$

for some $\alpha, \beta \geq 0$, where $\mathcal{C}_1, ..., \mathcal{C}_\alpha$ are the components of size greater than 2, and $\mathcal{C}_{\alpha+1}, ..., \mathcal{C}_{\alpha+\beta}$ denote the components of size 2. Finally, $\mathcal{D} = \{D_1, ..., D_s\}$ denotes the set of isolated vertices (that are connected by the edges in $\mathcal{G}^{\neq}$).

For each component, we arbitrarily choose a *representative* $V_i \in \mathcal{C}_i$. When we assign a value to V_i, each vertex $W \in \mathcal{C}_i$ is automatically assigned the value $V_i \oplus \lambda(V_i, W)$ to satisfy the system of equations $\Gamma^=$. With the representatives, we define $\lambda_i(W) \stackrel{\text{def}}{=} \lambda(V_i, W)$ for simplicity. Any assignment to the representatives $(V_1, ..., V_{\alpha+\beta})$ makes all equations in the system $\Gamma^=$ be satisfied. Still, the assignment may not satisfy one of the conditions that

1. the assignments to $\mathcal{P}$ are different, and
2. some non-equations from $\Gamma^{\neq}$.

We also need to assign some values to the vertices in $\mathcal{D}$. Below, we clarify when the assignment satisfies the conditions, which can be written in terms of the non-equations.

Non-equations in the graph. Recall that $\mathcal{P}^{*2}$ denotes the set of pairs of different vertices included in the same set. We write $\mathcal{E}^{\neq}_{i,j} \subset \mathcal{C}_i \times \mathcal{C}_j$ for $i \neq j$[3] to denote the set of non-equations connecting vertices in $\mathcal{C}_i$ and $\mathcal{C}_j$.

We first consider that the assignments of $\mathcal{P}$ should be different. Fix arbitrary assignments of the representatives. Consider two vertices $(W, W') \in \mathcal{P}^{*2}$ such that $W \in \mathcal{C}_i$ and $W' \in \mathcal{C}_j$. If $i = j$, W and W' take different values due to the non-degeneracy regardless of the assignments of the representatives. For $i \neq j$, the condition $W \neq W'$ implies the non-equation

$$V_i \oplus \lambda_i(W) \neq V_j \oplus \lambda_j(W'). \tag{5}$$

Now we consider the edges in $\mathcal{E}^{\neq}$ with respect to Equation (4). Let $V \oplus V' \neq \mu$ be a non-equation in $\Gamma^{\neq}$ for $(V, V') \in \mathcal{E}^{\neq}_{i,j}$. For $\nu := \mu \oplus \lambda_i(V) \oplus \lambda_j(V')$, this non-equation can be written as

$$V_i \neq V_j \oplus \nu.$$

[3] We assume that $\mathcal{E}^{\neq}$ does not contain an edge connecting two vertices in the same component, which trivially holds or induces a contradiction.

If there is $(W, W') \in \mathcal{C}_i \times \mathcal{C}_j$ such that $\nu = \lambda_i(W) \oplus \lambda_j(W')$ holds, then we say the non-equation $V \oplus V' \neq \mu$ is *trivial*, because it can be derived from Equation (5). Also, if two non-equations in $\mathcal{E}_{i,j}^{\neq}$ give the same ν, we say that they are *equivalent*. We assume that $\Gamma^{\neq}$ does not include trivial non-equations or equivalent non-equation pairs.

Let $c_i := |\mathcal{C}_i|$ be the number of vertices in $\mathcal{C}_i$ and $v_{i,j} = |\mathcal{E}_{i,j}^{\neq}|$ be the number of $\neq$-labeled edges connecting a vertex in $\mathcal{C}_i$ and a vertex in $\mathcal{C}_j$. We write $\mathcal{N}_{i,j}$ to denote the set of constants representing the non-equations between V_i and V_j for $i \neq j$:

$$\{\lambda_i(W) \oplus \lambda_j(W')\}_{(W,W') \in \mathcal{E}_{i,j}} \cup \{\mu \oplus \lambda_i(V) \oplus \lambda_j(V')\}_{(V,V') \in \mathcal{T}_{i,j} : [V \oplus V' \neq \mu] \in \Gamma^{\neq}} ,$$

where the assignments of V_i and V_j must obey the condition $V_i \notin \mathcal{N}_{i,j} \oplus V_j$. Note that the size of $\mathcal{N}_{i,j}$ is computed by $c_i c_j + v_{i,j}$ because we assume that the graph does not have trivial or equivalent non-equations. Define a set $\mathcal{N}_i := \cup_{j<i} \mathcal{N}_{i,j}$,

We say that Γ (and $\mathcal{G}(\Gamma)$) is *nice* if $\mathcal{G}^{=}$ is a non-degenerated acyclic bipartite graph, and for any $(\lambda, \neq)$-labeled edge between (P, Q), there is no λ-labeled trail between P and Q in $\mathcal{G}^{=}$.

Counting the number of solutions. For the system Γ with its associated graph $\mathcal{G} = \mathcal{G}(\Gamma)$, we write the set of the solutions, or the valid assignments to $\{V_1, ..., V_{\alpha+\beta}\} \cup \mathcal{D}$, of $\mathcal{G}$ by $\mathcal{S}(\mathcal{G})$, and denote the number of solutions by $h(\mathcal{G}) = |\mathcal{S}(\mathcal{G})|$. We use the following notations in the analysis.

- For a set $I \subset [\alpha + \beta]$, $\mathcal{S}_I$ denotes the set of partial assignments to $\{V_i\}_{i \in I}$ that satisfying all the conditions, or *solutions*, and $h_I := |\mathcal{S}_I|$ be the number of solutions for $\{V_i\}_{i \in I}$. If $I = [i]$ for some $i \leq \alpha + \beta$, we simply use $\mathcal{S}_i$ and h_i instead of $\mathcal{S}_I$ and h_I, respectively.
- Recall that $v_{i,j}$ denotes the number of $\neq$-labeled edges between $\mathcal{C}_i$ and $\mathcal{C}_j$. Let v_i be the number of $\neq$-labeled edges connecting a vertex in $\mathcal{C}_i$ and $\mathcal{C}_j$ for some $j < i$, so that $v_i = \sum_{j<i} v_{i,j}$. Let $v_{j,I}$ be the number of $\neq$-labeled edges connecting $\mathcal{C}_j$ and $\mathcal{C}_i$ for some $i \in I$. For the set $\mathcal{N}_{i,j}$ of constants representing the non-equations between V_i, V_j, define $\mathcal{N}_{i,j}(V_j) = \mathcal{N}_{i,j} \oplus V_j$.
- For a set $I \subset [\alpha + \beta]$, we write $\mathcal{C}_I$ to denote the set of vertices $\cup_{i \in I} \mathcal{C}_i$. The number of vertices are denoted by $c_i = |\mathcal{C}_i|$ and $C_I = |\mathcal{C}_I|$. When $I = [i]$, we simply write C_i instead of C_I. Let $\xi_{\max} := \max_i \{c_i\}$.

We also define the following sets for $i \in [\alpha + \beta]$:

$$\mathcal{R}_i \stackrel{\text{def}}{=} \left\{(V_1, V_1', V_2, V_2') \in {\mathcal{C}_i}^{*2} \times {\mathcal{C}_j}^{*2} \,\middle|\, j < i \text{ and } \lambda(V_1, V_1') = \lambda(V_2, V_2')\right\}. \quad (6)$$

Theorem 3 (Mirror Theory for $\xi_{max} > 2$). *Let $\mathcal{G}$ ge a nice graph, let q denote the number of edges of $\mathcal{G}$, and q_c denote the number of edges of $\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_\alpha$. When $q \leq \frac{N}{4\xi_{\max}}$ and $0 < q_c \leq q$, it holds that*

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{|\mathcal{V}|}} - 1\right| \leq \exp\left(\frac{18v + 2\sum_{i=1}^{\alpha+\beta} |\mathcal{R}_i| + 2\sum_{i=1}^{\alpha} c_i^2}{N} + \frac{31 q_c q^2 + 2q_c^2 \sum_{i=1}^{\alpha} c_i^2}{N^2} + \frac{20q^4}{N^3}\right) - 1.$$

In particular, if

$$\frac{18v + 2\sum_{i=1}^{\alpha+\beta}|\mathcal{R}_i| + 2\sum_{i=1}^{\alpha}c_i^2}{N} + \frac{31q_cq^2 + 2q_c^2\sum_{i=1}^{\alpha}c_i^2}{N^2} + \frac{20q^4}{N^3} \le 1$$

we have

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{|\mathcal{V}|}} - 1\right| \le \frac{36v + 4\sum_{i=1}^{\alpha+\beta}|\mathcal{R}_i| + 4\sum_{i=1}^{\alpha}c_i^2}{N} + \frac{62q_cq^2 + 4q_c^2\sum_{i=1}^{\alpha}c_i^2}{N^2} + \frac{40q^4}{N^3}.$$

This theorem combines Theorem 5 and Theorem 7 — the lower and upper bounds of

$$\frac{h(\mathcal{G})(N-1)^q}{(N)_{|\mathcal{V}|}}$$

where the statements and proofs are in Sections 4.1 and 4.3 and the final statement is from $e^x \le 1 + 2x$ for $x \le 1$.

Below, we give a Mirror theory for equations systems with all component sizes 2. Theorem 4 is used in a multi-user security proof of XoP.

Theorem 4 (Mirror Theory with $\xi_{max} = 2$). *Let $q \le \frac{N}{13}$ and $q_c = 0$. Then, it holds*

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} - 1\right| \le \exp\left(\frac{3\sum_{i=1}^q|\mathcal{R}_i|}{N} + \frac{3q^2}{N^2} + \frac{10(n+1)^2}{N}\right) - 1.$$

Further, if $\frac{3\sum_{i=1}^q|\mathcal{R}_i|}{N} + \frac{3q^2}{N^2} + \frac{10(n+1)^2}{N} < 1$, it holds

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} - 1\right| \le \frac{6\sum_{i=1}^q|\mathcal{R}_i|}{N} + \frac{6q^2}{N^2} + \frac{20(n+1)^2}{N}.$$

Proof. By Theorem 8 (see Section 4.4 for details), we have

$$\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} - 1 \le \exp\left(\frac{3\sum_{i=1}^q|\mathcal{R}_i|}{N} + \frac{147q^3}{N^3} + \frac{10(n+1)^2}{N}\right) - 1,$$

and by Theorem 6 (deferred to the end of this section), we have

$$\begin{aligned} 1 - \frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} &\le \frac{2q^2}{N^2} + \frac{128q^3}{N^3} + \frac{8(n+1)^3}{3N^2} \\ &\le \exp\left(\frac{2q^2}{N^2} + \frac{128q^3}{N^3} + \frac{8(n+1)^3}{3N^2}\right) - 1. \end{aligned}$$

Since $\frac{128q^3}{N^3} \le \frac{147q^3}{N^3} \le \frac{2q^2}{N^2}$, and $\frac{8(n+1)^3}{3N^2} \le \frac{10(n+1)^2}{N}$, we conclude with

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} - 1\right| \le \exp\left(\frac{3\sum_{i=1}^q|\mathcal{R}_i|}{N} + \frac{3q^2}{N^2} + \frac{10(n+1)^2}{N}\right) - 1.$$

The last statement can be proved using the fact $\exp(X) - 1 \le 2X$ for $X < 1$. $\square$

Theorems 5 and 6 are Mirror theory lower bounds for equations systems with all component sizes being 2, and with all component sizes larger or equal to 2, separately. They are used in the multi-user security proof of DbHtS discussed in Section 7. Specifically, Theorem 6 is used in the proof of Theorem 12 and Theorem 5 is used in the proof of Theorem 13.

Theorem 5 (Lower Bound Mirror Theory for $\xi_{max} > 2$). *Assume that $8q \leq N$. It holds that*

$$\frac{h(\mathcal{G})(N-1)^q}{(N)_{|\mathcal{V}|}} \geq 1 - \frac{9q_c^2 \sum_{1\leq i\leq \alpha} c_i^2}{8N^2} - \frac{31q_c q^2}{N^2} - \frac{16q^4}{N^3} - \frac{18v}{N}.$$

The proof of this theorem is deferred to Section 4.1.

Theorem 6 (Lower Bound Mirror Theory for $\xi_{max} = 2$). *Let $q \leq \frac{N}{13}$ and $q_c = 0$. Then, it holds that*

$$\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} \geq 1 - \frac{2q^2}{N^2} - \frac{128q^3}{N^3} - \frac{8(n+1)^3}{3N^2}.$$

The proof of this theorem is deferred to Section 4.2.

4.1 Proof of Mirror Theory - Lower Bound for $\xi_{max} > 2$

Below, we describe the proof of Theorem 5, which is a Mirror theory lower bound for equations systems with all component sizes larger or equal to 2. Many parts of the proof are adapted from [16] while we modified some parts for our purpose. The following simple bounds of $h_{I\cup\{j\}}$ in terms of h_I for $j \notin I$ will be useful.

Lemma 7. *Recall h_I for $I \subset [\alpha+\beta]$ is the number of the valid assignments of $\{V_i\}_{i\in I}$. For $I \subsetneq [\alpha+\beta]$ and $j \in [\alpha+\beta] \setminus I$, it holds that*

$$(N - c_j C_I - v_{j,I})h_I \leq h_{I\cup\{j\}} \leq N h_I.$$

In particular, the following inequality holds

$$(N - c_{i+1}C_i - v_{i+1})h_i \leq h_{i+1} \leq N h_i.$$

Proof. The upper bound is clear because V_j can take one of $[N]$ values. For the lower bound, fix an assignment $V_I = \{V_i\}_{i\in I} \in \mathcal{S}_I$. The assignment to V_j cannot take the values in $\cup_{i\in I}\mathcal{N}_{i,j}(V_i)$. By the union bound, the size of this set is bounded above by $\sum_i c_i c_j + v_{j,I} = C_I c_j + v_{j,I}$, and V_j can take at least $(N - c_j C_I - v_{j,I})$ different values for each solution V_I. □

Components of size > 2. The following lemma shows a rudimentary Mirror lower bound of $h(\mathcal{G})$ for the components of size > 2. Let $v^{(\geq 3)} = \sum_{i=1}^{\alpha} v_i$.

Lemma 8. *It holds that*

$$\frac{h_\alpha(N-1)^{q_c}}{(N)_{C_\alpha}} \geq 1 - \frac{C_\alpha^2 \sum_{1\leq i\leq \alpha} c_i^2}{N^2} - \frac{2v^{(\geq 3)}}{N}.$$

Proof. We first prove the following claim.

Claim. For each $0 \le i < \alpha$ such that $c_{i+1}C_i \le N$, it holds that

$$\frac{h_{i+1}(N-1)^{c_{i+1}-1}}{h_i(N-C_i)_{c_{i+1}}} \ge 1 - \left(\frac{c_{i+1}C_i}{N}\right)^2 - \frac{2v_{i+1}}{N}.$$

Proof (of claim). By applying Lemma 7 to h_{i+1}, we obtain

$$\frac{h_{i+1}(N-1)^{c_{i+1}-1}}{h_i(N-C_i)_{c_{i+1}}} \ge \frac{N - c_{i+1}C_i - v_{i+1}}{N} \cdot \frac{N(N-1)^{c_{i+1}-1}}{(N-C_i)_{c_{i+1}}}.$$

The second term is bounded below by

$$\begin{aligned}\frac{N(N-1)^{c_{i+1}-1}}{(N-C_i)_{c_{i+1}}} &= \left(1 + \frac{C_i}{N-C_i}\right) \cdot \left(1 + \frac{C_i}{N-C_i-1}\right)^{c_{i+1}-1} \\ &\ge \left(1 + \frac{C_i}{N}\right)^{c_{i+1}} \ge 1 + \frac{c_{i+1}C_i}{N},\end{aligned}$$

which gives the overall lower bound $\left(1 - \frac{c_{i+1}C_i}{N} - \frac{v_{i+1}}{N}\right) \cdot \left(1 + \frac{c_{i+1}C_i}{N}\right) \ge 1 - \left(\frac{c_{i+1}C_i}{N}\right)^2 - \frac{2v_{i+1}}{N}$ as we wanted. □

Now we return to the original proof. If there exists i such that $c_{i+1}C_i \ge N$, the right-hand side is less than 0 as follows so that the inequality becomes obvious:

$$(C_\alpha c_{i+1})^2 \ge (C_i c_{i+1})^2 \ge N^2.$$

When $c_{i+1}C_i \le N$ holds for all $i \le \alpha - 1$, we obtain the desired result by multiplying the inequalities from the claim for $i = 1, ..., \alpha-1$ and using Inequality (2), and the fact that $C_i \le C_\alpha$ for $i \le \alpha$. □

Components of size 2. The following lemma is for the components of size 2. Let $v^{(2)} = \sum_{i=\alpha+1}^{\alpha+\beta} v_i$.

Lemma 9. *Suppose that $4C_{\alpha+\beta} + 2 \le N$. Then it holds that*

$$\frac{h_{\alpha+\beta}(N-1)^\beta}{h_\alpha(N-C_\alpha)_{2\beta}} \ge 1 - \frac{4C_\alpha^2\beta}{N^2} - \frac{4C_\alpha\beta^2}{N^2} - \frac{22\beta^2}{N^2} - \frac{32C_\alpha\beta^3}{3N^3} - \frac{16\beta^4}{N^3} - \frac{18v^{(2)}}{N}.$$

Proof. We use the following claim.

Claim. For each $0 \le i < \beta$ such that $4C_{\alpha+i} + 2 \le N$, it holds that

$$\frac{h_{\alpha+i+1}(N-1)}{h_{\alpha+i}(N-C_{\alpha+i})_2} \ge 1 - \frac{4C_\alpha^2}{N^2} - \frac{8C_\alpha i}{N^2} - \frac{44i}{N^2} - \frac{32C_\alpha i^2}{N^3} - \frac{64i^3}{N^3} - \frac{2v_{i+1}}{N} - \frac{16v^{(2)}}{N^2}.$$

Proof (of claim). We adapt the inequality bottom of [16, page 14].

$$
\begin{aligned}
&\frac{h_{\alpha+i+1}(N-1)}{h_{\alpha+i}(N-C_{\alpha+i})_2} \geq \frac{(N-1)\left(N-2C_{\alpha+i}-v_{\alpha+i+1}+\frac{4i^2-16i-8v^{(2)}}{N}(1-\frac{4C_{\alpha+i}}{N})\right)}{(N-C_{\alpha+i})_2} \\
&\geq \frac{N^2-(2C_{\alpha+i}+1)N-v_{\alpha+i+1}N+(4i^2-16i-8v)(1-\frac{4C_{\alpha+i}+1}{N})}{N^2-(2C_{\alpha+i}+1)N+C_{\alpha+i}(C_{\alpha+i}+1)} \\
&= 1-\frac{v_{\alpha+i+1}N+C_{\alpha+i}(C_{\alpha+i}+1)-(4i^2-16i-8v^{(2)})(1-\frac{4C_{\alpha+i}+1}{N})}{N^2-(2C_{\alpha+i}+1)N+C_{\alpha+i}(C_{\alpha+i}+1)} \\
&\geq 1-\frac{4C_\alpha^2}{N^2}-\frac{8C_\alpha i}{N^2}-\frac{36i}{N^2}-\frac{32C_\alpha i^2}{N^3}-\frac{64i^3}{N^3}-\frac{8i^2}{N^3}-\frac{2v_{\alpha+i+1}}{N}-\frac{16v^{(2)}}{N^2} \\
&\geq 1-\frac{4C_\alpha^2}{N^2}-\frac{8C_\alpha i}{N^2}-\frac{44i}{N^2}-\frac{32C_\alpha i^2}{N^3}-\frac{64i^3}{N^3}-\frac{2v_{\alpha+i+1}}{N}-\frac{16v^{(2)}}{N^2}
\end{aligned}
$$

The first inequality is adapted from [16, Bottom of page 14], and the second inequality uses $N-1 \leq N$ (in the third term) and $(1-x)(1-y) \geq 1-x-y$ for $x = 1/N$ and $y = 4C_{\alpha+i}/N$ (in the last term). In the third inequality, we use that the denominator is less than $N^2/2$ because $2C_{\alpha+i}+1 \leq N/2$. The last inequality removes some non-dominating terms. $\square$

By multiplying the above inequality for $i = 0, ..., \beta-1$, we have:

$$
\begin{aligned}
&\frac{h_{\alpha+\beta}(N-1)^\beta}{h_\alpha(N-C_\alpha)_{2\beta}} = \prod_{i=0}^{\beta-1}\frac{h_{\alpha+i+1}(N-1)}{h_{\alpha+i}(N-C_{\alpha+i})_2} \\
&\geq \prod_{i=0}^{\beta-1}\left(1-\frac{4C_\alpha^2}{N^2}-\frac{8C_\alpha i}{N^2}-\frac{44i}{N^2}-\frac{32C_\alpha i^2}{N^3}-\frac{64i^3}{N^3}-\frac{2v_{i+1}}{N}-\frac{16v^{(2)}}{N^2}\right) \\
&\geq 1-\sum_{i=0}^{\beta-1}\left(\frac{4C_\alpha^2}{N^2}+\frac{8C_\alpha i}{N^2}+\frac{44i}{N^2}+\frac{32C_\alpha i^2}{N^3}+\frac{64i^3}{N^3}+\frac{2v_{i+1}}{N}+\frac{16v}{N^2}\right) \\
&\geq 1-\frac{4C_\alpha^2\beta}{N^2}-\frac{4C_\alpha\beta^2}{N^2}-\frac{22\beta^2}{N^2}-\frac{32C_\alpha\beta^3}{3N^3}-\frac{16\beta^4}{N^3}-\frac{18v^{(2)}}{N}.
\end{aligned}
$$

The first inequality is from the claim, and the second one is Inequality (2). In the last inequality, we use Inequality (3) and $\beta \leq N$. $\square$

Isolated vertices. Finally, we need to exclude the solutions that violate some non-equations connected to $\mathcal{D}$. Let $v_{\mathcal{D}}$ be the number of such non-equations.

Lemma 10. *Suppose that $C_{\alpha+\beta}+|\mathcal{D}| \leq N/2$. It holds that*

$$\frac{h(\mathcal{G})}{h_{\alpha+\beta}(N-C_{\alpha+\beta})_{|\mathcal{D}|}} \geq 1-\frac{2v_{\mathcal{D}}}{N}.$$

Proof. For each solution to $\mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup ... \sqcup \mathcal{C}_{\alpha+\beta}$, there is $(N-C_{\alpha+\beta})_{|\mathcal{D}|}$ valid assignments to the vertices in $\mathcal{D}$ ignoring the non-equations. Among them, at most $(N - C_{\alpha+\beta})_{|\mathcal{D}|-1}$ assignments violate each non-equation. Therefore we have

$$h(\mathcal{G}) \geq h_{\alpha+\beta} \cdot \left((N - C_{\alpha+\beta})_{|\mathcal{D}|} - v_{\mathcal{D}}(N - C_{\alpha+\beta})_{|\mathcal{D}|-1}\right),$$

and the desired inequality follows from the condition $C_{\alpha+\beta} + |\mathcal{D}| \leq N/2$. $\square$

Proof of Theorem 5. Observe that $8q \leq N$ implies the conditions of all lemmas. Applying Lemmas 8 to 10 in sequence, we have

$$\begin{aligned}
&\frac{h(\mathcal{G})(N-1)^{q_c+\beta}}{(N)_{C_\alpha+\beta+|\mathcal{D}|}} \geq 1 - \frac{C_\alpha^2 \sum_{1\leq i\leq \alpha} c_i^2}{N^2} - \frac{4C_\alpha^2\beta + 4C_\alpha\beta^2 + 22\beta^2}{N^2} \\
&- \frac{32C_\alpha\beta^3/3 + 16\beta^4}{N^3} - \frac{2v^{(\geq 3)} + 18v^{(2)} + 2v_{\mathcal{D}}}{N} \\
&\geq 1 - \frac{9q_c^2 \sum_{1\leq i\leq \alpha} c_i^2}{4N^2} - \frac{9q_c^2\beta + 6q_c\beta^2 + 22\beta^2}{N^2} - \frac{16q_c\beta^3 + 16\beta^4}{N^3} - \frac{18v}{N} \\
&\geq 1 - \frac{9q_c^2 \sum_{1\leq i\leq \alpha} c_i^2}{8N^2} - \frac{31q_cq^2}{N^2} - \frac{16q^4}{N^3} - \frac{18v}{N}.
\end{aligned}$$

We use $3q_c \geq 2C_\alpha$, and $v = v^{(\geq 3)} + v^{(2)} + v_{\mathcal{D}}$ in the first inequality. In the second inequality, we use $q_c + \beta = q$ so that $\beta \leq q$ and $\sum_{i=1}^{\alpha} c_i = C_\alpha \leq 3q_c/2$, and finally $q_c \geq 1$ to suppress the β^2 term. $\square$

4.2 Proof of Mirror Theory - Lower Bound for $\xi_{max} = 2$

The following concepts and useful auxiliary lemma compute the more refined lower bound for mirror theory with $\xi_{\max} = 2$ —Theorem 6.

For $m \in \{2, \cdots, q\}$, let $\mathcal{I} = \{i_1, \cdots, i_m\} \subset [q]$ be an index set such that $|\mathcal{I}| = m$. We define

$$\begin{aligned}
\mathcal{V}[\mathcal{I}] &\stackrel{\text{def}}{=} \{P_{\gamma_{i_1}}, P_{\gamma'_{i_1}}, \cdots, P_{\gamma_{i_m}}, P_{\gamma'_{i_m}}\}, \\
\mathcal{E}[\mathcal{I}] &\stackrel{\text{def}}{=} \{(P_{\gamma_{i_1}}, P_{\gamma'_{i_1}}, \lambda_{i_1}), \cdots, (P_{\gamma_{i_m}}, P_{\gamma'_{i_m}}, \lambda_{i_m})\}, \\
\mathcal{G}[\mathcal{I}] &\stackrel{\text{def}}{=} (\mathcal{V}[\mathcal{I}], \mathcal{E}[\mathcal{I}]),
\end{aligned}$$

where $(P_\gamma, P_{\gamma'}, \lambda) \in \mathcal{E}[\mathcal{I}]$ represents an edge connecting P_γ and $P_{\gamma'}$ with label λ. When $\mathcal{I} = [m]$, we will simply write $\mathcal{G}_m$ to denote $\mathcal{G}[\mathcal{I}]$. So $\mathcal{G}_q = \mathcal{G}(\Gamma)$, which is the graph representation of the equation system Γ. We also define

$$\mathcal{R}[\mathcal{I}]_i \stackrel{\text{def}}{=} \left\{(V_1, V_1', V_2, V_2') \in {\mathcal{C}_i}^{*2} \times {\mathcal{C}_j}^{*2} \,\middle|\, i, j \in \mathcal{I} \text{ and } j < i \text{ and } \lambda(V_1, V_1') = \lambda(V_2, V_2')\right\}.$$

For $k \in [m-1]$, let $\mathcal{J} = (j_1, j_2, \cdots, j_{k+1}) \in \mathcal{I}^{k+1}$ be a sequence of *distinct* indices in $\mathcal{I}$, and let $\mathcal{L} = (L_1, \cdots, L_k) \in (\{0,1\}^n \setminus \{0\}^n)^k$ be a sequence of n-bit weights. Then we define an edge set (a set of equations) $\mathcal{F}[\mathcal{J}, \mathcal{L}] \stackrel{\text{def}}{=}$

$\{(P_{\gamma_{j_1}}, P_{\gamma'_{j_2}}, L_1), \cdots, (P_{\gamma_{j_k}}, P_{\gamma'_{j_{k+1}}}, L_k)\}$ and a weighted graph (an equation system) $\mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}] \stackrel{\text{def}}{=} \mathcal{G}[\mathcal{I}] \cup \mathcal{F}[\mathcal{J}, \mathcal{L}]$. When $h(\mathcal{G}[\mathcal{I}] \cup \mathcal{F}[\mathcal{J}, \mathcal{L}]) > 0$, we say that $\mathcal{G}[\mathcal{I}] \cup \mathcal{F}[\mathcal{J}, \mathcal{L}]$ is valid. We also define subgraphs of $\mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}]$ as follows

$$\mathcal{G}^{-+}[\mathcal{I}, \mathcal{J}, \mathcal{L}] \stackrel{\text{def}}{=} \mathcal{G}[\mathcal{I}] \cup (\mathcal{F}[\mathcal{J}, \mathcal{L}] \setminus \{(P_{\gamma_{j_1}}, P_{\gamma'_{j_2}}, L_1)\}),$$
$$\mathcal{G}^{+-}[\mathcal{I}, \mathcal{J}, \mathcal{L}] \stackrel{\text{def}}{=} \mathcal{G}[\mathcal{I} \setminus \{j_{k+1}\}] \cup (\mathcal{F}[\mathcal{J}, \mathcal{L}] \setminus \{(P_{\gamma_{j_k}}, P_{\gamma'_{j_{k+1}}}, L_k)\}),$$
$$\mathcal{G}^{--}[\mathcal{I}, \mathcal{J}, \mathcal{L}] \stackrel{\text{def}}{=} \mathcal{G}[\mathcal{I} \setminus \{j_{k+1}\}] \cup (\mathcal{F}[\mathcal{J}, \mathcal{L}] \setminus \{(P_{\gamma_{j_1}}, P_{\gamma'_{j_2}}, L_1), (P_{\gamma_{j_k}}, P_{\gamma'_{j_{k+1}}}, L_k)\}).$$

When $\mathcal{I}, \mathcal{J}, \mathcal{L}$ are clear from the context, we will simply write

$$\mathcal{G}^{++} = \mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}], \mathcal{G}^{-+} = \mathcal{G}^{-+}[\mathcal{I}, \mathcal{J}, \mathcal{L}], \mathcal{G}^{+-} = \mathcal{G}^{+-}[\mathcal{I}, \mathcal{J}, \mathcal{L}], \mathcal{G}^{--} = \mathcal{G}^{--}[\mathcal{I}, \mathcal{J}, \mathcal{L}].$$

Lemma 11 (Orange Equation). *Let $\alpha = 0$. For any positive integer $t \in \{1, \cdots, q\}$, it holds*

$$h_t = (N - 2C_{t-1} + |\mathcal{R}_t|)h_{t-1} + \sum_{E \in \mathbb{L}[\mathcal{G}_t]} h(\mathcal{G}_{t-1} \cup E),$$

where $\mathbb{L}[\mathcal{G}_t] = \{(V, V', \lambda_t) | 0 \le i < j < t, V \in C_i, V' \in C_j, h(\mathcal{G}_{t-1} \cup (V, V', \lambda_t)) > 0\}$.

Proof. For $t = 1, \cdots q$, recall the component $\mathcal{C}_t$ has only two vertices and one edge and λ_t be the label of the edge in $\mathcal{C}_t$. Define the set $\Lambda_t \stackrel{\text{def}}{=} (\bigsqcup_{i \in [t]} \mathcal{C}_i)$. We thus have

$$\begin{aligned}
h_t &= \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} \Big(N - |\Lambda_{t-1} \bigcup (\Lambda_{t-1} \oplus \lambda_t)|\Big) \\
&= \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} \Big(N - |\Lambda_{t-1}| - |\Lambda_{t-1} \oplus \lambda_t| + |\Lambda_{t-1} \bigcap (\Lambda_{t-1} \oplus \lambda_t)|\Big) \\
&= (N - 2C_{t-1})h_{t-1} + \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} |\Lambda_{t-1} \bigcap (\Lambda_{t-1} \oplus \lambda_t)|, \qquad (7)
\end{aligned}$$

where $\mathcal{S}_{t-1}$ is the set of solutions to $\mathcal{G}_{t-1}$. In particular, we have

$$\sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} |\Lambda_{t-1} \bigcap (\Lambda_{t-1} \oplus \lambda_t)| = \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} \sum_{V, V' \in \Lambda_{t-1}} \mathbb{1}(V \oplus V' = \lambda_t).$$

Let us consider following cases for a fixed pair of (V, V'):

1. For each $(W, W', V, V') \in \mathcal{R}_t$, we have

$$\sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} \mathbb{1}(V \oplus V' = \lambda_t) = \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} 1 = h_{t-1}.$$

2. If $V \in C_i, V' \in C_j, i < j < t$, then we have

$$\sum_{(V_1,\dots,V_{t-1})\in\mathcal{S}_{t-1}} \mathbb{1}(V \oplus V' = \lambda_t) = h(\mathcal{G}_{t-1} \cup \{(V, V', \lambda_t)\}).$$

This leads to

$$\begin{aligned}
&\sum_{(V_1,\dots,V_{t-1})\in\mathcal{S}_{t-1}} \sum_{V,V'\in\Lambda_{t-1}} \mathbb{1}(V \oplus V' = \lambda_t) \\
=& \sum_{(V_1,\dots,V_{t-1})\in\mathcal{S}_{t-1}} \left(\sum_{(W,W',V,V')\in\mathcal{R}_t} \mathbb{1}(V \oplus V' = \lambda_t) + \sum_{V\in C_i, V'\in C_j, i<j<t} \mathbb{1}(V \oplus V' = \lambda_t) \right) \\
=& \; |\mathcal{R}_t|\, h_{t-1} + \sum_{E\in\mathcal{L}[\mathcal{G}_t]} h(\mathcal{G}_{t-1} \cup E). \qquad (8)
\end{aligned}$$

Lemma 11 follows from Equations (7) and (8). □

Lemma 12 (Purple Equation). *Let $\alpha = 0$. Fix integers m, k such that $1 \leq k < m \leq q$, an index set $\mathcal{I} \subset [q]$ such that $|\mathcal{I}| = m$, a sequence of distinct indices $\mathcal{J} = (j_1, \cdots, j_{k+1}) \in \mathcal{I}^{k+1}$ and a sequence of labels $\mathcal{L} = (L_1, \cdots, L_k) \in (\{0,1\} \setminus \{0^n\})^k$. If $\mathcal{G}^{++}(\mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}])$ is valid, then it holds*

$$h(\mathcal{G}^{++}) = h(\mathcal{G}^{+-}) - \sum_{E\in\mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\}) + \sum_{\{E,E'\}\in\mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E, E'\}),$$

where

$$\begin{aligned}
\mathbb{M}[\mathcal{G}^{++}] = &\{E = (P_{\gamma_{j_k}}, V, L_k \oplus \lambda_{j_{k+1}} \oplus \lambda_a) : V' \in \mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}], V, V' \in \mathcal{C}_a, h(\mathcal{G}^{+-} \cup \{E\}) > 0\} \\
&\cup \{E = (P_{\gamma_{j_k}}, V, L_k \oplus \lambda_a) : V' \in \mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}], V, V' \in \mathcal{C}_a, h(\mathcal{G}^{+-} \cup \{E\}) > 0\} \\
\mathbb{N}[\mathcal{G}^{++}] = &\{\{E, E'\} = \{(P_{\gamma_{j_k}}, V, L_k \oplus \lambda_{j_{k+1}} \oplus \lambda_a), (V', W, \lambda_{j_{k+1}}\} : \\
&W, V' \in \mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}], W \neq V', V, V' \in \mathcal{C}_a, h(\mathcal{G}^{+-} \cup \{E, E'\}) > 0\}.
\end{aligned}$$

Proof. Without loss of generality, we assume that $\mathcal{I} = [m]$, $\mathcal{J} = \{m-k, m-k+1, \cdots, m\}$. Let $\mathcal{S} \subset (\{0,1\}^n)^{2m}$ and $\mathcal{S}' \subset (\{0,1\}^n)^{2m-2}$ be the sets of solutions to $\mathcal{G}^{++}$ and $\mathcal{G}^{+-}$, respectively. For each solution $(P_{\gamma_1}, P_{\gamma'_1}, \cdots, P_{\gamma_{m-1}}, P_{\gamma'_{m-1}}) \in \mathcal{S}'$, let

$$\begin{aligned}
P_{\gamma_m} &= P_{\gamma_{m-1}} \oplus L_k \oplus \lambda_m \\
P_{\gamma'_m} &= P_{\gamma_{m-1}} \oplus L_k.
\end{aligned}$$

Then $(P_{\gamma_1}, P_{\gamma'_1}, \cdots, P_{\gamma_m}, P_{\gamma'_m})$ is a solution to $\mathcal{G}^{++}$ if and only all $2m$ variables have distinct values. Formally, it requires for any vertex $V \in \Lambda_{m-1}$,

$$\begin{aligned}
&P_{\gamma_m} \neq V \Leftrightarrow P_{\gamma_{m-1}} \oplus L_k \oplus \lambda_m \neq V \Leftrightarrow P_{\gamma_{m-1}} \neq V \oplus L_k \oplus \lambda_m \\
&P_{\gamma'_m} \neq V \Leftrightarrow P_{\gamma_{m-1}} \oplus L_k \neq V \Leftrightarrow P_{\gamma_{m-1}} \neq V \oplus L_k.
\end{aligned}$$

Therefore we have

$$\begin{aligned}
h(\mathcal{G}^{++}) = & \sum_{S\in\mathcal{S}'}(1-\mathbb{1}(P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k)\cup(\Lambda_{m-1}\oplus L_k\oplus\lambda_m))) \\
& = h(\mathcal{G}^{+-}) - \sum_{S\in\mathcal{S}'}\mathbb{1}(P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k)) - \sum_{S\in\mathcal{S}'}\mathbb{1}(P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k\oplus\lambda_m)) \\
& \quad + \sum_{S\in\mathcal{S}'}\mathbb{1}(P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k)\cap(\Lambda_{m-1}\oplus L_k\oplus\lambda_m))).
\end{aligned}$$

When $P_{\gamma_{m-1}} \in (\Lambda_{m-1}\oplus L_k\oplus\lambda_m)$, we know that the vertex $V = P_{\gamma_m}$ must satisfy $V\in\Lambda_{m-1}\setminus\mathcal{V}[\mathcal{J}]$. Otherwise there exists a trail such that $\lambda(V,P_{\gamma_m})=0$ in $\mathcal{G}^{++}$, which means $\mathcal{G}^{++}$ has a circle, invalid, a contradiction. Therefore this solution to $\mathcal{G}^{+-}$ is also a solution to $\mathcal{G}^{+-}\cup\{(P_{\gamma_{m-1}},V',L_k\oplus\lambda_m\oplus\lambda(V,V'))\}$, where V,V' are in the same component $\mathcal{C}$. Similarly, when $P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k)$, we know there exists a vertex $V=P_{\gamma_m}$ must satisfy $V\in\Lambda_{m-1}\setminus\mathcal{V}[\mathcal{J}]$, and this solution to $\mathcal{G}^{+-}$ is also a solution to $\mathcal{G}^{+-}\cup\{(P_{\gamma_{m-1}},V',L_k\oplus\lambda(V,V'))\}$, where V and V' are in the same component $\mathcal{C}$.

To summarize, we have

$$\begin{aligned}
\sum_{S\in\mathcal{S}'}\mathbb{1}(P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k\oplus\lambda_m)) &= \sum_{E\in\mathbb{M}_1} h(\mathcal{G}^{+-}\cup\{E\}),\\
\sum_{S\in\mathcal{S}'}\mathbb{1}(P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k)) &= \sum_{E\in\mathbb{M}_2} h(\mathcal{G}^{+-}\cup\{E\}),
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{M}_1 &\stackrel{\text{def}}{=} \{(P_{\gamma_{m-1}},V,L_k\oplus\lambda_m\oplus\lambda_a): V'\in\Lambda_{m-1}\setminus\mathcal{V}[\mathcal{J}], V,V'\in\mathcal{C}_a\},\\
\mathbb{M}_2 &\stackrel{\text{def}}{=} \{(P_{\gamma_{m-1}},V,L_k\oplus\lambda_a): V'\in\Lambda_{m-1}\setminus\mathcal{V}[\mathcal{J}], V,V'\in\mathcal{C}_a\}.
\end{aligned}$$

When $P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k)\cap(\Lambda_{m-1}\oplus L_k\oplus\lambda_m)$, we know there exists two distinct vertecies $V',W\in\Lambda_{m-1}\setminus\mathcal{V}[\mathcal{J}]$ such that $P_{\gamma_{m-1}} = V'\oplus L_k\oplus\lambda_m = W\oplus L_k$. Equivalently, for V such that $V\in\mathcal{C}_a$, we have $P_{\gamma_{m-1}} = V\oplus L_k\oplus\lambda_m\oplus\lambda_a = V'\oplus L_k\oplus\lambda_m = W\oplus L_k$. And this solution to $\mathcal{G}^{+-}$ is also a solution to $\mathcal{G}^{+-}\cup\{(P_{\gamma_{m-1}},V,L_k\oplus\lambda_m\oplus\lambda_a),(V',W,\lambda_m)\}$. Therefore, we have

$$\sum_{S\in\mathcal{S}'}\mathbb{1}(P_{\gamma_{m-1}}\in(\Lambda_{m-1}\oplus L_k)\cap(\Lambda_{m-1}\oplus L_k\oplus\lambda_m))) = \sum_{\{E,E'\}\in\mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-}\cup\{E,E'\}),$$

where

$$\mathbb{N}[\mathcal{G}^{++}] \stackrel{\text{def}}{=} \{\{(P_{\gamma_{m-1}},V,L_k\oplus\lambda_m\oplus\lambda_a),(V',W,\lambda_m)\} : W,V'\in\Lambda_{m-1}\setminus\mathcal{V}[\mathcal{J}], W\neq V', V,V'\in\mathcal{C}_a\}.$$

This concludes the proof. □

Lemma 13 estimates the size of sets $\mathbb{L}[\mathcal{G}_m]$, $\mathbb{M}[\mathcal{G}^{++}]$, and $\mathbb{N}[\mathcal{G}^{++}]$ using in Lemma 11 and 12. In order to state Lemma 13, we need to reorder the indices of $\mathcal{G}_q$; note that any reordering of the indices does not affect the number of solutions to $\mathcal{G}_q$. For the edge set $\{(P_{\gamma_1}, P_{\gamma'_1}, \lambda_1), \cdots, (P_{\gamma_q}, P_{\gamma'_q}, \lambda_q)\}$, we choose as many different label λ as possible, put them in a separate list, remove them from the edge set, and perform the same procedure recursively for the remaining elements. This procedure defines a reordering of the edges (indices) and with it, we have

$$\max_{\lambda\in\{0,1\}^n\setminus\{0^n\}} \{|\{k \leq m : \lambda_k = \lambda\}|\} \leq |\mathcal{R}_{m+1}|. \tag{9}$$

Lemma 13 (Size Lemma). *Fix integer m, k, n such that $2 \leq k < m \leq t \leq q$. Then, it holds that*

$$|\mathbb{L}[\mathcal{G}_m]| = (m - 1 - |\mathcal{R}_m|)(m - 2 - |\mathcal{R}_m|).$$

For an index set $\mathcal{I} \subset [t]$ such that $|\mathcal{I}| = m$, a sequence of distinct indices $\mathcal{J} = (j_1, \cdots, j_{k+1}) \in \mathcal{I}^{k+1}$ and a sequence of labels $\mathcal{L} = (L_1, \cdots, L_k) \in (\{0,1\} \setminus \{0^n\})^k$. If $\mathcal{G}^{++}(\mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}])$ is valid, then it holds

$$\begin{aligned} \left|\mathbb{M}[\mathcal{G}^{-+}]\right| - 4(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{M}[\mathcal{G}^{++}]\right| \leq 2r, \\ \left|\mathbb{N}[\mathcal{G}^{-+}]\right| - 4r(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{N}[\mathcal{G}^{++}]\right| \leq r^2. \end{aligned}$$

When $k = 1$, it holds

$$\begin{aligned} 2m - |\mathcal{R}[\mathcal{I}]_m| - 4(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{M}[\mathcal{G}^{++}]\right| \leq 2r, \\ \left|\mathbb{L}[\mathcal{G}^{-+}]\right| - 4r(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{N}[\mathcal{G}^{++}]\right| \leq r^2. \end{aligned}$$

Proof. 1. For the first equality, we first recall the definition of $\mathbb{L}[\mathcal{G}_i] = \{(V, V', \lambda_m) | 0 \leq j_1 < j_2 < m, V \in C_{j_1}, V' \in C_{j_2}, h(\mathcal{G}_{i-1} \cup (V, V', \lambda_m)) > 0\}$. Since $\lambda_{j_1} \neq \lambda_m$ and $\lambda_{j_2} \neq \lambda_m$ otherwise the resulting graph is invalid. The number of such edge is

$$(m - 1 - |\mathcal{R}_m|)(m - 1 - |\mathcal{R}_m| - 1), \tag{10}$$

which proves the statement.

2. We then prove the second inequality. Note that $\mathbb{M}[\mathcal{G}^{++}] \subset \mathbb{M}[\mathcal{G}^{-+}]$ when $k \geq 2$. We consider the edge in $\mathbb{M}[\mathcal{G}^{-+}] \setminus \mathbb{M}[\mathcal{G}^{++}]$, which is of the form either $(P_{\gamma_{j_k}}, V, L_k \oplus \lambda_{j_{k+1}} \oplus \lambda_a)$ or $(P_{\gamma_{j_k}}, V, L_k \oplus \lambda_a)$ for $V' \in (\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]) \cup \mathcal{C}_{j_1}$ and $V, V' \in \mathcal{C}_a$. Such an edge falls into at least one of the following three cases.
   (a) $V \in \mathcal{C}_{j_1}$. At most four edges fall into this case since $|\mathcal{C}_{j_1}| = 2$ and V has at most two possible assigned values.
   (b) $E = (P_{\gamma_{j_k}}, V, L_k \oplus \lambda_{j_{k+1}} \oplus \lambda_a)$ for $V' \in \mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]$ and $V, V' \in \mathcal{C}_a$. Since $E \in \mathbb{M}[\mathcal{G}^{-+}] \setminus \mathbb{M}[\mathcal{G}^{++}]$, by $\mathbb{M}$'s definition, we know $\mathcal{G}^{++}$ and $\mathcal{G}^{--} \cup \{E\}$ are valid, while $\mathcal{G}^{+-} \cup \{E\}$ is invalid. This means $\lambda(V, P_{\gamma_{j_1}}) = 0$ or $\lambda(V, P_{\gamma'_{j_1}}) = 0$. For the case $\lambda(V, P_{\gamma_{j_1}}) = 0$, we have

$$\lambda_a = L_1 \oplus \cdots \oplus L_k \oplus \lambda_{\gamma_{j_2}} \oplus \cdots \oplus \lambda_{\gamma_{j_{k+1}}} (\overset{\text{def}}{=} \lambda).$$

The number of such edges E is at most $|\{a \leq t : \lambda_a = \lambda\}|$ where by Equation 9,

$$|\{a \leq t : \lambda_a = \lambda\}| \leq |\mathcal{R}_{t+1}|\,.$$

Similarly, the number of edges satisfying $\lambda(V, P_{\gamma'_{j_1}}) = 0$ is at most $|\mathcal{R}_{t+1}|\,.$

(c) $E = (P_{\gamma_{j_k}}, V, L_k \oplus \lambda_a)$ for $V' \in \mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]$ and $V, V' \in \mathcal{C}_a$. Similarly to Case 2, the total number of edges of this type is at most $2\,|\mathcal{R}_{t+1}|$.

Moreover, $|\mathbb{M}[\mathcal{G}^{++}]| \leq 2r$. We conclude that

$$\left|\mathbb{M}[\mathcal{G}^{-+}]\right| - 4(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{M}[\mathcal{G}^{++}]\right| \leq 2r. \tag{11}$$

3. We then prove the third inequality. Note that $\mathbb{N}[\mathcal{G}^{++}] \subset \mathbb{N}[\mathcal{G}^{-+}]$ when $k \geq 2$. We consider the pair of edges in $\mathbb{N}[\mathcal{G}^{-+}] \setminus \mathbb{N}[\mathcal{G}^{++}]$, where the edge E is of the form $(P_{\gamma_{j_k}}, V, L_k \oplus \lambda_{j_{k+1}} \oplus \lambda_a)$ for $V' \in (\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]) \cup \mathcal{C}_{j_1}$ and $V, V' \in \mathcal{C}_a$ and the edge E' is of the form $(V', W, \lambda_{j_{k+1}})$ for $W \in (\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]) \cup \mathcal{C}_{j_1}$, $W \neq V'$. Such a pair $\{E, E'\}$ falls into at least one of the following three cases.
   (a) $V' \in (\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]) \cup \mathcal{C}_{j_1}$ and $W \in \mathcal{C}_{j_1}$. Since $|(\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]) \cup \mathcal{C}_{j_1}| \leq r$, the number of pairs of edges is at most $2r$.
   (b) $W \in (\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]) \cup \mathcal{C}_{j_1}$ and $V' \in \mathcal{C}_{j_1}$. Similarly to case 1, the number of such pairs of edges is at most $2r$.
   (c) $V', W \in (\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}])$. By $\mathbb{N}$'s definition, we know $\mathcal{G}^{++}$ and $\mathcal{G}^{--} \cup \{E, E'\}$ are valid, while $\mathcal{G}^{+-} \cup \{E, E'\}$ is invalid. This means $\lambda(V, P_{\gamma_{j_1}}) = 0$ or $\lambda(V, P_{\gamma'_{j_1}}) = 0$ or $\lambda(W, P_{\gamma_{j_1}}) = 0$ or $\lambda(W, P_{\gamma'_{j_1}}) = 0$. For the case $\lambda(V, P_{\gamma_{j_1}}) = 0$, we have

$$\lambda_a = L_1 \oplus \cdots \oplus L_k \oplus \lambda_{\gamma_{j_2}} \oplus \cdots \oplus \lambda_{\gamma_{j_{k+1}}} (\overset{\text{def}}{=} \lambda).$$

   The number of such edges E is at most $|\{a \leq t : \lambda_a = \lambda\}|$ where by Equation 9

$$|\{a \leq t : \lambda_a = \lambda\}| \leq |\mathcal{R}_{t+1}|\,.$$

   So the number of edge pair $\{E, E'\}$ of this type is at most $|\mathcal{R}_{t+1}|\, r$. The number of edge pairs for the other three cases follows the same upper bound.

   Moreover, $|\mathbb{N}[\mathcal{G}^{++}]| \leq r^2$. We conclude that

$$\left|\mathbb{N}[\mathcal{G}^{-+}]\right| - 4r(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{N}[\mathcal{G}^{++}]\right| \leq r^2. \tag{12}$$

4. We then turn to the fourth inequality. When $k = 1$, we define the edge set $\mathbb{M}'$ whose edge is of the form either $(P_{\gamma_{j_1}}, V, L_1 \oplus \lambda_{j_2} \oplus \lambda_a)$ or $(P_{\gamma_{j_1}}, V, L_1 \oplus \lambda_a)$ for $V' \in (\mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}]) \cup \mathcal{C}_{j_1}$ and $V, V' \in \mathcal{C}_a$. Note that $|\mathbb{M}'| = 2m - |\mathcal{R}[\mathcal{I}]_{j_2}| \geq 2m - |\mathcal{R}[\mathcal{I}]_m|$ and $\mathbb{M}[\mathcal{G}^{++}] \subset \mathbb{M}'$. We then follow a similar analysis procedure as that in the proof of the second inequality and can conclude that

$$2m - |\mathcal{R}[\mathcal{I}]_m| - 4(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{M}[\mathcal{G}^{++}]\right| \leq 2r. \tag{13}$$

5. We finally turn to the fifth inequality. When $k = 1$, we define the pairs of edges set $\mathbb{N}'$ where $E = (P_{\gamma_{j_1}}, V, L_1 \oplus \lambda_{j_2} \oplus \lambda_a)$ and $E' = (V', W, \lambda_{j_2})$ such that $W, V' \in \mathcal{V}[\mathcal{I}] \setminus \mathcal{V}[\mathcal{J}], W \neq V', V, V' \in \mathcal{C}_a, h(\mathcal{G}^{+-} \cup \{E'\}) > 0$. Then we have $\mathbb{N}[\mathcal{G}^{++}] \subset \mathbb{N}'$ and $|\mathbb{N}'| = |\mathbb{L}[\mathcal{G}^{-+}]|$ since $\mathbb{L}[\mathcal{G}^{-+}]$ is obtained by collecting E' for all $\{E, E'\} \in \mathbb{N}'$. We then follow a similar analysis procedure as that in the proof of the third inequality and can conclude that

$$\left|\mathbb{L}[\mathcal{G}^{-+}]\right| - 4r(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{N}[\mathcal{G}^{++}]\right| \leq r^2. \tag{14}$$

By Equation (10) and inequalities (11) to (14), the proof is completed. □

The following combinatorial lemma proved by [17] is used in our Mirror theory statement.

Lemma 14. *Let t be a positive integer, and let $(D_{m,k})_{m,k}$ be a two-dimensional sequence of non-negative numbers, where $1 \leq m \leq t$ and $k \leq m-1$. If $D_{m,k} = 0$ for $k \leq 0$, and*

$$D_{m,k} \leq D_{m-1,k-1} + 2A \cdot D_{m-1,k} + A^2 \cdot D_{m-1,k+1} + \frac{C}{(N-2A)^{t-m+k}},$$

for $2 \leq m \leq t$ and $k \leq m-3$, where A, C are positive constants and $A < 2^{n-1}$. Then, for any integer c such that $1 \leq c \leq \frac{m}{2} - 1$, it holds

$$D_{m,1} \leq \sum_{i=c}^{2c} \binom{2c}{i} A^i D_{m-c,1-c+i} + \sum_{j=0}^{c-1} \sum_{i=j}^{2j} \binom{2j}{i} \frac{A^i C}{(N-2A)^{t-m+1+i}}.$$

We define the following two-dimensional sequence $D^t_{m,k}$ where t is a fixed positive integer such that $t \leq q$, $1 \leq m \leq t$ and k is an integer

- When $1 \leq k \leq m-1$,

$$D^t_{m,k} = \max_{\mathcal{I},\mathcal{J},\mathcal{L}} \left\{ \left| \frac{h(\mathcal{G}^{-+}[\mathcal{I},\mathcal{J},\mathcal{L}])}{N} - h(\mathcal{G}[\mathcal{I},\mathcal{J},\mathcal{L}]) \right| \right\},$$

  where the maximum is taken over all possible index sets $\mathcal{I} \subset [t]$ such that $|\mathcal{I}| = m$, sequence of distinct indices $\mathcal{J} \in \mathcal{I}^{k+1}$, and sequence of labels $\mathcal{L} \in (\{0,1\}^n \setminus \{0^n\})^k$ such that $\mathcal{G}[\mathcal{I},\mathcal{J},\mathcal{L}]$ is valid.
- When $k \leq 0$, $D^t_{m,k} = 0$.

In order to upper bound $D^t_{m,k}$, we begin with the following lemma.

Lemma 15. *For any $\mathcal{I} \subset [t], \mathcal{J} \in \mathcal{I}^{k+1}, \mathcal{L} \in (\{0,1\}^n \setminus \{0^n\})^k$ such that $|\mathcal{I}| = m$ and $\mathcal{G}[\mathcal{I},\mathcal{J},\mathcal{L}]$ is valid, one has*

$$h(\mathcal{G}[\mathcal{I},\mathcal{J},\mathcal{L}]) \leq \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+k}}.$$

Proof. Without loss of generality, let $\mathcal{I} = [m]$ and $\mathcal{S}$ be the set of solution to $\mathcal{G}_m$. For each solution $(P_{\gamma_1}, P_{\gamma'_1}, \cdots, P_{\gamma_m}, P_{\gamma'_m}) \in \mathcal{S}$, $(P_{\gamma_1}, P_{\gamma'_1}, \cdots, P_{\gamma_{m+1}}, P_{\gamma'_{m+1}})$ is a solution to $\mathcal{G}_{m+1}$ if for all $V \in \varLambda_m$, $P_{\gamma_{m+1}} \neq V, P_{\gamma'_{m+1}} \neq V$. Therefore, we have

$$\begin{aligned} h(\mathcal{G}_{m+1}) &\geq \sum_{S \in \mathcal{S}} (N - \left| \{V \in \varLambda_m : V = P_{\gamma_{m+1}}\} \cup \{V \in \varLambda_m : V = P_{\gamma'_{m+1}}\} \right|) \\ &\geq (N - 2r) h(\mathcal{G}_m). \end{aligned}$$

By repeatedly applying the above inequality, we have

$$h(\mathcal{G}_m) \leq \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m}}, \tag{15}$$

which completes the statement when $k = 0$.

When $k \geq 1$, let $\mathcal{L} \in (\{0,1\}^n \setminus \{0^n\})^k$ and without loss of generality let $\mathcal{L} = \{m-k, m-k+1, \cdots, m\}$. For each solution $(P_{\gamma_1}, P_{\gamma'_1}, \cdots, P_{\gamma_m}, P_{\gamma'_m})$ to $\mathcal{G}^{++}$ (let its solution set be $\mathcal{S}'$), $(P_{\gamma_1}, P_{\gamma'_1}, \cdots, P_{\gamma_{m-k}}, P_{\gamma'_{m-k}}, \cdots, P_{\gamma_m}, P_{\gamma'_m})$ is a solution to $\mathcal{G}^{-+}$ if for all $V \in \varLambda_m \setminus \mathcal{C}_{m-k}$, $P_{\gamma_{m-k}} \neq V, P_{\gamma'_{m-k}} \neq V$. Therefore, we have

$$\begin{aligned} h(\mathcal{G}^{-+}) &\geq \sum_{S \in \mathcal{S}'} (N - \left| \{V \in \varLambda_m \setminus \mathcal{C}_{m-k} : V = P_{\gamma_{m-k}}\} \cup \{V \in \varLambda_m \setminus \mathcal{C}_{m-k} : V = P_{\gamma'_{m-k}}\} \right|) \\ &\geq (N - 2r) h(\mathcal{G}^{++}). \end{aligned}$$

By repeatedly applying the above inequality, we have

$$h(\mathcal{G}^{++}) \leq \frac{h(\mathcal{G}_m)}{(N-2r)^{k}}. \tag{16}$$

Combining Equation (15) and (16), we complete the proof. $\square$

By lemma 15, for $\mathcal{G}^{++} (= \mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}])$,

$$\frac{h(\mathcal{G}^{-+})}{N} \leq \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+k-1}} \leq \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+k}}.$$

Therefore, using the above inequality and $D^t_{m,k}$'s definition,

$$D^t_{m,k} \leq \max\{\frac{h(\mathcal{G}^{-+})}{N}, h(\mathcal{G}^{++})\} \leq \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+k}}. \tag{17}$$

Lemma 16 shows that our constructed two-dimensional sequence $D^t_{m,k}$ satisfies the condition required for using combinatorial Lemma 14. This proof is based on purple equation (Lemma 12), size lemma (Lemma 13) and Lemma 15.

Lemma 16. *For $2 \leq m \leq t$, and $k \leq m-3$, it holds that*

$$D^t_{m,k} \leq D^t_{m-1,k-1} + 2r \cdot D^t_{m-1,k} + r^2 \cdot D^t_{m-1,k+1} + \frac{C}{(N-2r)^{t-m+k}},$$

where

$$C \stackrel{\text{def}}{=} \frac{(6\,|\mathcal{R}_{t+1}| + 6) h(\mathcal{G}_t)}{N}.$$

Proof. When $m = 2$ or 3, it is easy to see the statement holds since by $D^t_{m,k}$'s definition $D^t_{m,k} = 0$ when $k \leq 0$.

When $m \geq 4$ and $2 \leq k \leq m-3$, for any $\mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}]$ such that $|\mathcal{I}| = m, \mathcal{J} \in \mathcal{I}^{k+1}$, and $\mathcal{L} \in (\{0,1\}^n \setminus \{0^n\})^k$, by purple equation (Lemma 12), we have

$$h(\mathcal{G}^{++}) = h(\mathcal{G}^{+-}) - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\}) + \sum_{\{E,E'\} \in \mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E, E'\}), \tag{18}$$

$$h(\mathcal{G}^{-+}) = h(\mathcal{G}^{--}) - \sum_{E \in \mathbb{M}[\mathcal{G}^{-+}]} h(\mathcal{G}^{--} \cup \{E\}) + \sum_{\{E,E'\} \in \mathbb{N}[\mathcal{G}^{-+}]} h(\mathcal{G}^{--} \cup \{E, E'\}). \tag{19}$$

By $D^t_{m,k}$'s definition and since $\mathcal{G}^{--} = (\mathcal{G}^{+-})^{-+}$, we have

$$\left| \frac{h(\mathcal{G}^{--})}{N} - h(\mathcal{G}^{+-}) \right| \leq D^t_{m-1,k-1}.$$

For each edge $E \in \mathbb{M}[\mathcal{G}^{++}]$, by $D^t_{m,k}$'s definition, we have

$$\left| \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} - h(\mathcal{G}^{+-} \cup \{E\}) \right| \leq D^t_{m-1,k}.$$

Using the above inequality, we have

$$\begin{aligned}
&\left| \sum_{E \in \mathbb{M}[\mathcal{G}^{-+}]} \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\}) \right| \\
&\leq \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} \left| \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} - h(\mathcal{G}^{+-} \cup \{E\}) \right| + \sum_{E \in \mathbb{M}[\mathcal{G}^{-+}] \setminus \mathbb{M}[\mathcal{G}^{++}]} \left| \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} \right| \\
&\leq 2r \cdot D^t_{m-1,k} + 4(|\mathcal{R}_{t+1}| + 1) \left| \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} \right| \quad \text{(by size lemma (Lemma 13))} \\
&\leq 2r \cdot D^t_{m-1,k} + \frac{4(|\mathcal{R}_{t+1}| + 1) h(\mathcal{G}_t)}{N(N-2r)^{t-m+k}}. \qquad \text{(by Lemma 15)}
\end{aligned}$$

For each pair of edge $\{E, E'\} \in \mathbb{N}[\mathcal{G}^{++}]$, since $\mathcal{G}^{--} \cup \{E, E'\} = (\mathcal{G}^{+-} \cup \{E, E'\})^{-+}$, we have

$$\left| \frac{h(\mathcal{G}^{--} \cup \{E, E'\})}{N} - h(\mathcal{G}^{+-} \cup \{E, E'\}) \right| \leq D^t_{m-1,k+1}.$$

Using the above inequality, Lemma 13 and 15, we have

$$\begin{aligned}
&\left| \sum_{E \in \mathbb{N}[\mathcal{G}^{-+}]} \frac{h(\mathcal{G}^{--} \cup \{E, E'\})}{N} - \sum_{E \in \mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E, E'\}) \right| \\
&\leq r^2 D^t_{m-1,k+1} + \frac{4r(|\mathcal{R}_{t+1}| + 1) h(\mathcal{G}_t)}{N(N-2r)^{t-m+k+1}}.
\end{aligned}$$

By subtracting Equation (18) from $\frac{1}{N}\times$ Equation (19) and combine everything above, we have

$$
\begin{aligned}
&\left|\frac{h(\mathcal{G}^{-+})}{N} - h(\mathcal{G}^{++})\right| \\
&\le D^t_{m-1,k-1} + 2r \cdot D^t_{m-1,k} + \frac{4(|\mathcal{R}_{t+1}|+1)h(\mathcal{G}_t)}{N(N-2r)^{t-m+k}} + r^2 \cdot D^t_{m-1,k+1} + \frac{4r(|\mathcal{R}_{t+1}|+1)h(\mathcal{G}_t)}{N(N-2r)^{t-m+k+1}} \\
&\le D^t_{m-1,k-1} + 2r \cdot D^t_{m-1,k} + r^2 \cdot D^t_{m-1,k+1} + \frac{(6\,|\mathcal{R}_{t+1}|+6)h(\mathcal{G}_t)}{N(N-2r)^{t-m+k}}. \\
&\hfill (\because \tfrac{2r}{N-2r} \le 1)
\end{aligned}
$$

When $m \ge 4$ and $k = m-3 = 1$, for any $\mathcal{G}[\mathcal{I},\mathcal{J},\mathcal{L}]$ such that $|\mathcal{I}| = m, \mathcal{J} = j_1, j_2 \in \mathcal{I}^2$ and $\mathcal{L} \in \{0,1\}^n \setminus \{0^n\}$. By Purple equation (Lemma 12) and Orange equation (Lemma 11), respectively, we have

$$
h(\mathcal{G}^{++}) = h(\mathcal{G}^{+-}) - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\}) + \sum_{\{E,E'\} \in \mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E,E'\}), \tag{20}
$$

$$
h(\mathcal{G}^{-+}) = h(\mathcal{G}^{--}) - (2m-2-|\mathcal{R}[\mathcal{I}]_m|)h(\mathcal{G}^{--}) + \sum_{E \in \mathcal{L}[\mathcal{G}^{-+}]} h(\mathcal{G}^{--} \cup E). \tag{21}
$$

Since $\mathcal{G}^{+-} = \mathcal{G}^{--}$, we have $h(\mathcal{G}^{--}) - h(\mathcal{G}^{+-}) = 0$. For each edge $E \in \mathbb{M}[\mathcal{G}^{++}]$, by $D^t_{m,k}$'s definition, we have

$$
\left|\frac{h(\mathcal{G}^{--})}{N} - h(\mathcal{G}^{+-} \cup \{E\})\right| \le D^t_{m-1,1}.
$$

Using the above inequality, we have

$$
\begin{aligned}
&\left|(2m-2-|\mathcal{R}[\mathcal{I}]_m|)\frac{h(\mathcal{G}^{--})}{N} - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\})\right| \\
&\le \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} \left|\frac{h(\mathcal{G}^{--})}{N} - h(\mathcal{G}^{+-} \cup \{E\})\right| + \big|2m-2-|\mathcal{R}[\mathcal{I}]_m| - \left|\mathbb{M}[\mathcal{G}^{++}]\right|\big| \frac{h(\mathcal{G}^{--})}{N} \\
&\le 2r \cdot D^t_{m-1,1} + 4(|\mathcal{R}_{t+1}|+1)\frac{h(\mathcal{G}^{--})}{N} && \text{(by Lemma 13)} \\
&\le 2r \cdot D^t_{m-1,1} + \frac{4(|\mathcal{R}_{t+1}|+1)h(\mathcal{G}_t)}{N(N-2r)^{t-m+1}}.
\end{aligned}
$$

For each pair of edge $\{E,E'\} \in \mathbb{N}[\mathcal{G}^{++}]$, since each edge E uniquely determines an edge E', we have

$$
\left|\frac{h(\mathcal{G}^{--} \cup \{E\})}{N} - h(\mathcal{G}^{+-} \cup \{E,E'\})\right| \le D^t_{m-1,2}.
$$

It implies that

$$\left| \sum_{E \in \mathcal{L}[\mathcal{G}^{-+}]} \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} - \sum_{\{E,E'\} \in \mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E, E'\}) \right|$$
$$\leq r^2 \cdot D^t_{m-1,2} + \frac{4r(|\mathcal{R}_{t+1}| + 1)h(\mathcal{G}_t)}{N(N-2r)^{t-m+2}}.$$

By subtracting Equation (20) from $\frac{1}{N}\times$ Equation (21) and combining the above, we have

$$D^t_{m,1} = \max_{\mathcal{I},\mathcal{J},\mathcal{L}} \left| \frac{h(\mathcal{G}^{-+})}{N} - h(\mathcal{G}^{++}) \right|$$
$$\leq 2r \cdot D^t_{m-1,1} + r^2 \cdot D^t_{m-1,2} + \frac{(6\,|\mathcal{R}_{t+1}| + 6)h(\mathcal{G}_t)}{N(N-2r)^{t-m+1}}.$$

This concludes the proof. □

When $k = 1$, Lemma 17 gives a sharper upper bound on $D^t_{t,1}$. The proof can be derived from Lemma 16 and 14.

Lemma 17. *If $2n + 2 \leq t < q$ and $r \leq \frac{N}{13}$, then it holds that*

$$D^t_{t,1} \leq \frac{(29\,|\mathcal{R}_{t+1}| + 31)h(\mathcal{G}_t)}{N^2}.$$

Proof. Since the two-dimensional sequence $D^t_{m,k}$ satisfies Lemma 16, let $n \le \frac{m}{2} - 1$ (the c in Lemma 14), then we can apply Lemma 14 to obtain

$$\begin{aligned}
D^t_{m,1} &\le \sum_{i=n}^{2n} \binom{2n}{i} r^i D^t_{m-n,1-n+i} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j}\binom{2j}{i}\frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\le \sum_{i=n}^{2n} (2e)^i r^i D^t_{m-n,1-n+i} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j}\binom{2j}{i}\frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\qquad\qquad (\tbinom{2n}{i} \le (\tfrac{2ne}{i})^i \le (2e)^i \text{ when } n \le i \le 2n) \\
&\le \sum_{i=n}^{2n} (2er)^i \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+1+i}} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j}\binom{2j}{i}\frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\qquad\qquad (\text{by Inequality } 17) \\
&= \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \sum_{i=n}^{2n} (\frac{2er}{N-2r})^i + \sum_{j=0}^{n-1}\sum_{i=j}^{2j}\binom{2j}{i}\frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\le \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \sum_{i=n}^{\infty} (1/2)^i + \sum_{j=0}^{n-1}\sum_{i=j}^{2j}\binom{2j}{i}\frac{r^i C}{(N-2r)^{t-m+1+i}} \quad (r \le \tfrac{N}{13}) \\
&\le \frac{2h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \frac{1}{2^n} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j}\binom{2j}{i}\frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\le \frac{2h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \frac{1}{2^n} + \frac{4C}{(N-2r)^{t-m+1}}.
\end{aligned}$$

Now, plug in $m = t$ and have

$$\begin{aligned}
D^t_{t,1} &\le \frac{2h(\mathcal{G}_t)}{(N-2r)} \frac{1}{2^n} + \frac{4C}{(N-2r)} \\
&\le \frac{\frac{26}{11} h(\mathcal{G}_t)}{N^2} + \frac{\frac{13}{11}(24\,|\mathcal{R}_{t+1}| + 24)h(\mathcal{G}_t)}{N^2} \\
&\qquad (\because r \le \tfrac{N}{13}, \text{ substitute } C \text{ defined in Lemma } 16) \\
&= \frac{(29\,|\mathcal{R}_{t+1}| + 31)h(\mathcal{G}_t)}{N^2}.
\end{aligned}$$

This concludes the proof. □

Finally, using the above, we can prove Theorem 6 as follows:

Proof (of Theorem 6). Recall when $q_c = 0$, we have $\alpha = 0$. We also know in this equation system $C_i = 2i$, since each component has size only 2. To recursively compute the lower bound for $\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_\beta}}$, we first lower bound $\frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)}$ for $i = 0, \cdots, 2n+1$ and $i = 2n+2, \cdots, q$, separately.

To lower bound each term of $\frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)}$, we first lower bound h_{i+1} by h_i for $i = 0, \cdots, 2n+1$ and $i = 2n+2, \cdots, q-1$, separately. By lemma 7, we simply have $(N - c_{i+1}C_i)h_i \le h_{i+1}$ for $i = 0, \cdots, 2n+1$ as $v_{i+1} = 0$ in graph $\mathcal{G}$, which represents the equation system Γ.

For $i \ge 2n+2$, we first replicate part of the proof of Lemma 11 and have

$$h_{i+1} = (N - 2C_i)h_i + |\mathcal{R}_{i+1}|h_i + \sum_{\{V,V'\} \in \mathbb{L}_{i+1}} h'(V, V'), \tag{22}$$

where recall $h'(V, V')$ denote the number of solutions to Λ_i such that $V \oplus V' = \lambda_{i+1}$ for $V, V' \in \Lambda_i$, and $\mathbb{L}_{i+1} \stackrel{\text{def}}{=} \left\{\{V, V'\} \in \Lambda_i^{*2} \,\middle|\, \lambda(V, V') = \perp\right\}$. We also have $|\mathbb{L}_{i+1}| = C_i(C_i - 2) = 4i^2 - 4i$. Then by Lemma 17, we have

$$h'(V, V') \ge \frac{h_i}{N}\left(1 - \frac{(29\,|\mathcal{R}_{i+1}| + 31)}{N}\right).$$

Plugging in Equation (22), we have

$$h_{i+1} \ge \left(N - 4i + |\mathcal{R}_{i+1}| + \frac{4i^2 - 4i}{N} - \frac{116\,|\mathcal{R}_{i+1}|\,i^2 - 116\,|\mathcal{R}_{i+1}|\,i + 124i^2 - 124i}{N^2}\right) h_i.$$

For $i = 2n+2, \cdots, q$, plugging in the above inequality, we have

$$\begin{aligned}
\frac{h_{i+1}(N-1)}{h_i(N - C_i)(N - C_i - 1)} &\ge \frac{(N-1)\left(N - 4i + |\mathcal{R}_{i+1}| + \frac{4i^2 - 4i}{N} - \frac{116|\mathcal{R}_{i+1}|i^2 - 116|\mathcal{R}_{i+1}|i + 124i^2 - 124i}{N^2}\right)}{N^2 - (4i+1)N + 4i^2 + 2i} \\
&\ge \frac{N^2 - (4i+1)N + 4i^2 + \frac{128i - 128i^2}{N} + \frac{(N-1)N|\mathcal{R}_{i+1}| - 116i^2|\mathcal{R}_{i+1}|}{N}}{N^2 - (4i+1)N + 4i^2 + 2i} \\
&\ge 1 + \frac{-2i + \frac{128i - 128i^2}{N}}{N^2} \qquad\qquad (\because i \le q \le \tfrac{N}{13}) \\
&\ge 1 - \frac{2q}{N^2} - \frac{128q^2}{N^3}.
\end{aligned}$$

For $i = 1, \cdots, 2n+1$, with $(N - c_{i+1}C_i)h_i \le h_{i+1}$, we have

$$\begin{aligned}
\frac{h_{i+1}(N-1)}{h_i(N - C_i)(N - C_i - 1)} &\ge \frac{(N - c_{i+1}C_i)(N-1)}{(N - C_i)(N - C_i - 1)} \\
&= \frac{N^2 - (4i+1)N + 4i}{N^2 - (4i+1)N + 4i^2 + 2i} \\
&\ge 1 - \frac{4i^2}{N^2}.
\end{aligned}$$

By using the above inequalities, then we have

$$\begin{aligned}
\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_\beta}} &= \prod_{i=0}^{2n+1} \frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i+1)} \times \prod_{i=2n+2}^{q-1} \frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i+1)} \\
&\geq \prod_{i=0}^{2n+1} \left(1 - \frac{4i^2}{N^2}\right) \times \prod_{i=2n+2}^{q-1} \left(1 - \frac{2q}{N^2} - \frac{128q^2}{N^3}\right) \\
&\geq \left(1 - \frac{4n(n+1)(2n+1)}{6N^2}\right) \left(1 - \frac{2q^2}{N^2} - \frac{128q^3}{N^3}\right) \\
&\geq 1 - \frac{2q^2}{N^2} - \frac{128q^3}{N^3} - \frac{8(n+1)^3}{3N^2},
\end{aligned}$$

which completes the proof. □

4.3 Proof of Mirror Theory - Upper Bound for $\xi_{\max} > 2$

Theorem 7 (Upper Bound Mirror Theory for $\xi_{\max} > 2$). *Let $\mathcal{G}$ be a nice graph, q denote the number of edges of $\mathcal{G}$, and q_c denote the number of edges of $\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_\alpha$.*

When $q \leq \frac{N}{4\xi_{\max}}$ and $0 < q_c \leq q$, then it holds that

$$\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} \leq \exp\left(\frac{2\sum_{i=1}^{\alpha+\beta} |\mathcal{R}_i| + 2\sum_{i=1}^{\alpha} c_i^2}{N} + \frac{2q_c^2 \sum_{i=1}^{\alpha} c_i^2 + 4q_c q^2}{N^2} + \frac{20q^4}{N^3} \right).$$

The proof of Theorem 7 is deferred to the end of this section. Before proving it, we introduce essential lemmas first.

Lemma 18. *When $q \leq \frac{N}{4\xi_{\max}}$ and $0 < q_c \leq q$, for $i = 0, \cdots, \alpha - 1$, it holds that*

$$h_{i+1} \leq \left(N - c_{i+1}C_i + |\mathcal{R}_{i+1}| + \frac{2(c_{i+1})_2 q_c^2}{N}\right) h_i.$$

Proof. For a vertex $V \in \mathcal{C}_{i+1}$, denote the set $\mathit{\Lambda}_V = (\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i) \oplus \lambda_{i+1}(V)$. Recall that $\mathcal{S}_i$ is the set of solutions to $(V_1, \ldots, V_i)$. By Fixing $\mathcal{S}_i$ and assigning any value to $V^* \in \mathcal{C}_{i+1}$, the other unknowns in $\mathcal{C}_{i+1}$ are uniquely determined. Hence a solution to h_{i+1} after fixing $\mathcal{S}_i$ can be identified to choose a solution to V^* from

$$\{0,1\}^n \setminus \bigcup_{V \in \mathcal{C}_{i+1}} \mathit{\Lambda}_V.$$

We thus have an upper bound of h_{i+1} as follows:

$$
\begin{aligned}
&\sum_{(V_1,\ldots,V_i)\in\mathcal{S}_i}\left(N-|\bigcup_{V\in\mathcal{C}_{i+1}}\Lambda_V|\right) && \text{(count for every fixed solution in } \mathcal{S}_i\text{)}\\
&\le \sum_{(V_1,\ldots,V_i)\in\mathcal{S}_i}\left(N-\sum_{V\in\mathcal{C}_{i+1}}|\Lambda_V|+\sum_{V,V'\in\mathcal{C}_{i+1}}|\Lambda_V\cap\Lambda_{V'}|\right) && \text{(Lemma 4)}\\
&\le \sum_{(V_1,\ldots,V_i)\in\mathcal{S}_i}\left(N-c_{i+1}C_i+\sum_{V,V'\in\mathcal{C}_{i+1}}|\Lambda_V\cap\Lambda_{V'}|\right)\\
&= (N-c_{i+1}C_i)h_i+\sum_{(V_1,\ldots,V_i)\in\mathcal{S}_i}\ \sum_{V,V'\in\mathcal{C}_{i+1}}|\Lambda_V\cap\Lambda_{V'}|.
\end{aligned}
$$

For $V_1, V_1' \in \mathcal{C}_{i+1}, V_2, V_2' \in \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$, let $h'(V_1, V_1', V_2, V_2')$ denote the number of solutions to $\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$ such that $V_2 \oplus V_2' = \lambda_{i+1}(V_1) \oplus \lambda_{i+1}(V_1')$. Let

$$
\mathbb{L}_{i+1} \stackrel{\text{def}}{=} \left\{\{V_1, V_1'\}, \{V_2, V_2'\} \in {\mathcal{C}_{i+1}}^{*2} \times (\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i)^{*2} \,\middle|\, \lambda(V_2, V_2') = \perp\right\}.
$$

Then the summation $\sum_{(V_1,\ldots,V_i)\in\mathcal{S}_i}\sum_{V,V'\in C_{i+1}}|\Lambda_V\cap\Lambda_{V'}|$ can be computed by

$$
|\mathcal{R}_{i+1}|h_i + \sum_{(\{V_1,V_1'\},\{V_2,V_2'\})\in\mathbb{L}_{i+1}} h'(V_1, V_1', V_2, V_2').
$$

This is because the constant in $\Lambda_{V_1} \cap \Lambda_{V_1'}$ satisfies that there exists $V_2, V_2' \in \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$ such that $V_2 \oplus V_2' = \lambda_{i+1}(V_1) \oplus \lambda_{i+1}(V_1')$. We count the number of such constants by considering two cases: V_2, V_2' are in the same component (the first term) or not (the second term).

Let $h''(V, V')$ denote the number of solutions to $(\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i) \setminus (\mathcal{C}_V \sqcup \mathcal{C}_{V'})$ where $V \in \mathcal{C}_V$ and $V' \in \mathcal{C}_{V'}$. For $(\{V_1, V_1'\}, \{V_2, V_2'\}) \in \mathbb{L}_{i+1}$, we have:

$$
\begin{aligned}
h'(V_1, V_1', V_2, V_2') &\le N \cdot h''(V_2, V_2') && \text{(Upper bound of Lemma 7)}\\
&\le \frac{Nh_i}{(N-\xi_{\max}C_i)^2} && \text{(Lower bound of Lemma 7)}\\
&\le \frac{h_i}{N}\left(1+\frac{2N\xi_{\max}C_i}{(N-\xi_{\max}C_i)^2}\right)\\
&\le \frac{h_i}{N}\left(1+\frac{192\xi_{\max}q_c}{25N}\right)\\
&\le \frac{73h_i}{25N},
\end{aligned}
$$

where the last two steps are because $C_i \le \frac{3q_c}{2}$ and $q_c \le q \le \frac{N}{4\xi_{\max}}$. We also compute

$$
|\mathbb{L}_{i+1}| \le \binom{c_{i+1}}{2}\binom{C_i}{2} \le \frac{(c_{i+1})_2 C_i^2}{4} \le \frac{9(c_{i+1})_2 q_c^2}{16}.
$$

Combining all together, we have

$$h_{i+1} \leq \left(N - c_{i+1}C_i + |\mathcal{R}_{i+1}| + \frac{2(c_{i+1})_2 q_c^2}{N} \right) h_i.$$

This concludes the proof. $\square$

Lemma 19. *For $\alpha > 0$ and $i = \alpha, \cdots, \alpha + \beta - 1$, it holds that*

$$h_{i+1} \leq \left(N - 2C_i + |\mathcal{R}_{i+1}| + \frac{C_i^2}{N} + \frac{3q_c q}{N} + \frac{16q^3}{N^2} \right) h_i$$

Proof. For $i = \alpha, \cdots, \alpha + \beta - 1$, recall the component $\mathcal{C}_i$ has only two vertices and one edge. Let λ_{i+1} be the label of the edge in $\mathcal{C}_{i+1}$ for such i in the proof's context. Denote the set $\Lambda_i \stackrel{\text{def}}{=} \bigsqcup_{j \in [i]} \mathcal{C}_i$ for $i = \alpha, \cdots, \alpha + \beta - 1$. We thus have

$$\begin{aligned} h_{i+1} &= \sum_{(V_1,\ldots,V_i) \in \mathcal{S}_i} \left(N - |\Lambda_i \bigcup (\Lambda_i \oplus \lambda_{i+1})| \right) \\ &= \sum_{(V_1,\ldots,V_i) \in \mathcal{S}_i} \left(N - |\Lambda_i| - |\Lambda_i \oplus \lambda_{i+1}| + |\Lambda_i \bigcap (\Lambda_i \oplus \lambda_{i+1})| \right) \\ &= (N - 2C_i) h_i + \sum_{(V_1,\ldots,V_i) \in \mathcal{S}_i} |\Lambda_i \bigcap (\Lambda_i \oplus \lambda_{i+1})|. \end{aligned} \tag{23}$$

For $V, V' \in \Lambda_i$, let $h'(V, V')$ denote the number of solutions to Λ_i such that $V \oplus V' = \lambda_{i+1}$. Let

$$\mathbb{M}_{i+1} \stackrel{\text{def}}{=} \left\{ \{V, V'\} \in \Lambda_i^{*2} \,\middle|\, \lambda(V, V') = \perp \right\}.$$

Then we have

$$\sum_{(V_1,\ldots,V_i) \in \mathcal{S}_i} |\Lambda_i \bigcap (\Lambda_i \oplus \lambda_{i+1})| = |\mathcal{R}_{i+1}| h_i + \sum_{\{V,V'\} \in \mathbb{M}_{i+1}} h'(V, V'). \tag{24}$$

Let $h''(V, V')$ denote the number of solution to $\Lambda_i \setminus (\mathcal{C}_V \sqcup \mathcal{C}_{V'})$ where $V \in \mathcal{C}_V$ and $V' \in \mathcal{C}_{V'}$.

Suppose that $V \in \mathcal{C}_j, V' \in \mathcal{C}_k$ for $j, k \leq i$. Applying Lemma 7, we have

$$\begin{aligned} h'(V, V') &\leq N \cdot h''(V, V') \\ &\leq \frac{N h_i}{(N - c_j C_i)(N - c_k C_i)} \\ &= \frac{h_i}{N} \left(1 + \frac{c_j C_i}{N - c_j C_i} \right) \left(1 + \frac{c_k C_i}{N - c_k C_i} \right) \\ &\leq \frac{h_i}{N} \left(1 + \frac{2c_j C_i}{N - c_j C_i} + \frac{2c_k C_i}{N - c_k C_i} \right), \end{aligned}$$

where we used $c_jC_i \le \xi_{\max} q \le N/4$ and $(1+x)(1+y) \le 1+2(x+y)$ for $x, y \le 1$. Since $\mathcal{C}_j$ has c_j vertices, the term related to j is added at most c_jC_i times. By $c_jC_i \le N - c_jC_i$, it holds that

$$\frac{(c_jC_i)^2}{N - c_jC_i} \le c_jC_i, \text{ and } \frac{(c_jC_i)^2}{N - c_jC_i} \le \frac{2(c_jC_i)^2}{N}.$$

Summing up over all (V, V'), we have

$$\begin{aligned}
h_{i+1} &\le \left(N - 2C_i + |\mathcal{R}_{i+1}| + \frac{C_i^2}{N} + \frac{\sum_{j=1}^{i} 2(c_jC_i)^2}{N(N - c_jC_i)} \right) h_i \\
&\le \left(N - 2C_i + |\mathcal{R}_{i+1}| + \frac{C_i^2}{N} + \sum_{j=1}^{\alpha} \frac{2c_jC_i}{N} + \sum_{j=\alpha+1}^{i} \frac{4(c_jC_i)^2}{N^2} \right) h_i \\
&\le \left(N - 2C_i + |\mathcal{R}_{i+1}| + \frac{C_i^2}{N} + \frac{3q_cq}{N} + \frac{16q^3}{N^2} \right) h_i,
\end{aligned}$$

where we use $\sum_{i=1}^{\alpha} c_i = C_\alpha \le 3q_c/2$, $C_i \le q$, $i \le \beta \le q$ and $c_j = 2$ for all $j \ge \alpha + 1$ for proving the last inequality.

Now we prove the second part of the statement, when $\alpha = 0$, which means there are no components with a size larger than 2 in the graph. For $V, V' \in \mathbb{L}_{i+1}$, if $i \ge 2n + 2$, by Lemma 17, then we have

$$\left| \frac{h_i}{N} - h'(V, V') \right| \le \frac{(29\,|\mathcal{R}_{i+1}| + 31)h_i}{N^2},$$

equivalently, we have

$$h'(V, V') \le \frac{h_i}{N} \left(1 + \frac{29\,|\mathcal{R}_{i+1}| + 31}{N} \right)$$

Plugging in Equation (24), we have

$$\begin{aligned}
\sum_{(V_1,\ldots,V_i) \in \mathcal{S}_i} |\Lambda_i \bigcap (\Lambda_i \oplus \lambda_i)| &= |\mathcal{R}_{i+1}| h_i + \sum_{\{V,V'\} \in \mathbb{L}_{i+1}} h'(V, V') \\
&\le \left(|\mathcal{R}_{i+1}| + \frac{C_i^2}{N} \left(1 + \frac{29\,|\mathcal{R}_{i+1}| + 31}{N} \right) \right) h_i.
\end{aligned}$$

Plugging the above inequality into Equation (23), we have

$$h_{i+1} \le \left(N - 2C_i + |\mathcal{R}_{i+1}| \left(1 + \frac{116q^2}{N^2} \right) + \frac{C_i^2}{N} + \frac{124q^2}{N^2} \right) h_i.$$

This concludes the proof. $\square$

Using the above lemmas, we can prove Theorem 7 as follows.

Proof (of Theorem 7). We start by finding a relation between h_i and h_{i+1}. Lemma 7 has already shown $h_{i+1} \le N h_i$, while Lemmas 18 and 19 gives us a tighter upper bound of h_{i+1} using h_i, of which the proofs are deferred to the end of this section. We first observe that the non-equations only decrease the number of solutions. So, in the following, we only consider a system of equations $\Gamma = \Gamma^{=}$.

Note that $\xi_{\max} \ge 3$, hence $q \le \frac{N}{12}$ by the constraints $4q\xi_{\max} \le N$. To recursively compute the upper bound for $\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}}$, we first upper bound $\frac{h_{i+1}(N-1)^{c_{i+1}-1}}{h_i(N-C_i)_{c_{i+1}}}$, for $i = 0, \cdots, \alpha - 1$. To do so, we observe

$$(N - C_i)_{c_{i+1}} \ge N^{c_{i+1}-1}(N - c_{i+1}C_{i+1}), \tag{25}$$

which is simply because by dividing $N^{c_{i+1}}$ from both side it is true that

$$\left(1 - \frac{C_i}{N}\right) \times \cdots \times \left(1 - \frac{C_i + c_{i+1} - 1}{N}\right) \ge \left(1 - \frac{C_i + c_{i+1}}{N}\right)^{c_{i+1}} \ge \left(1 - \frac{c_{i+1}C_{i+1}}{N}\right).$$

So we have

$$\begin{aligned}
&\frac{h_{i+1}(N-1)^{c_{i+1}-1}}{h_i(N-C_i)_{c_{i+1}}} \\
&\le \frac{(N-1)^{c_{i+1}-1}\left(N - c_{i+1}C_i + |\mathcal{R}_{i+1}| + \frac{2(c_{i+1})_2 q_c^2}{N}\right)}{N^{c_{i+1}-1}(N - c_{i+1}C_{i+1})} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(by Lemma 18 and Equation (25))} \\
&\le 1 + \frac{c_{i+1}C_{i+1} - c_{i+1}C_i}{N - c_{i+1}C_{i+1}} + \frac{|\mathcal{R}_{i+1}|}{N - c_{i+1}C_{i+1}} + \frac{2(c_{i+1})_2 q_c^2}{N(N - c_{i+1}C_{i+1})} \\
&\le 1 + \frac{c_{i+1}^2}{N - c_{i+1}C_{i+1}} + \frac{|\mathcal{R}_{i+1}|}{N - c_{i+1}C_{i+1}} + \frac{2(c_{i+1})_2 q_c^2}{N(N - c_{i+1}C_{i+1})} \\
&\le 1 + \frac{2c_{i+1}^2}{N} + \frac{2|\mathcal{R}_{i+1}|}{N} + \frac{4(c_{i+1})_2 q_c^2}{N^2}. \qquad (c_{i+1}C_{i+1} \le \tfrac{N}{2} \text{ by } 4q\xi_{\max} \le N)
\end{aligned}$$

Now we can compute

$$\begin{aligned}
\frac{h(\mathcal{G}_\alpha)(N-1)^{q_c}}{(N)_{C_\alpha}} &= \prod_{i=0}^{\alpha-1}\left(\frac{h_{i+1}(N-1)^{c_{i+1}-1}}{h_i(N-C_i)_{c_{i+1}}}\right) \\
&\le \prod_{i=0}^{\alpha-1}\left(1 + \frac{2|\mathcal{R}_{i+1}|}{N} + \frac{2c_{i+1}^2}{N} + \frac{4(c_{i+1})_2 q_c^2}{N^2}\right) \\
&\le \exp\left(\frac{2\sum_{i=1}^{\alpha}|R_i| + 2(\sum_{i=1}^{\alpha} c_i^2)}{N} + \frac{2q_c^2(\sum_{i=1}^{\alpha} c_i^2)}{N^2}\right),
\end{aligned}$$

where we use $1 + x \le e^x$. On the other hand, for $i = \alpha, \cdots, \alpha + \beta - 1$,

$$\begin{aligned}
&\frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)} \\
&\le \frac{(N-1)\left(N - 2C_i + |\mathcal{R}_{i+1}| + \frac{C_i^2}{N} + \frac{3q_c q}{N} + \frac{16q^3}{N^2}\right)}{(N-C_i)(N-C_i-1)} && \text{(by Lemma 19)} \\
&\le \frac{N^2 - (2C_i+1)N + C_i^2 + |R_{i+1}|\,N + 3q_c q + 16\frac{q^3}{N}}{N^2 - (2C_i+1)N + C_i^2} \\
&\le 1 + \frac{|R_{i+1}|\,N + 3q_c q + 16\frac{q^3}{N}}{N^2 - (2C_i+1)N + C_i^2} \\
&\le 1 + \frac{6\,|R_{i+1}|}{5N} + \frac{18q_c q}{5N^2} + \frac{96q^3}{5N^3}. && (2(C_i+1) \le 2q \le \tfrac{N}{6})
\end{aligned}$$

Now we can compute

$$\begin{aligned}
&\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} \\
&= \prod_{i=0}^{\alpha+\beta-1} \left(\frac{h_{i+1}(N-1)^{c_{i+1}-1}}{h_i(N-C_i)_{c_{i+1}}}\right) \\
&= \frac{h(\mathcal{G}_\alpha)(N-1)^{q_c}}{(N)_{C_\alpha}} \prod_{i=\alpha}^{\alpha+\beta-1} \left(\frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)}\right) \\
&\le \exp(\delta_1) \prod_{i=\alpha}^{\alpha+\beta-1} \left(1 + \frac{6\,|R_{i+1}|}{5N} + \frac{18q_c q}{5N^2} + \frac{96q^3}{5N^3}\right) \\
&\le \exp(\delta_1) \exp\left(\frac{2\sum_{i=\alpha}^{\alpha+\beta} |R_{i+1}|}{N} + \frac{4q_c q^2}{N^2} + \frac{20q^4}{N^3}\right) \\
&\le \exp(\delta_1 + \delta_2),
\end{aligned}$$

for

$$\delta_1 = \frac{2\sum_{i=1}^{\alpha} |R_i| + 2(\sum_{i=1}^{\alpha} c_i^2)}{N} + \frac{2q_c^2(\sum_{i=1}^{\alpha} c_i^2)}{N^2}$$

and

$$\delta_2 = \frac{2\sum_{i=\alpha+1}^{\alpha+\beta} |\mathcal{R}_i|}{N} + \frac{4q_c q^2}{N^2} + \frac{20q^4}{N^3},$$

where we use $1+x \le e^x$, $\beta \le q$, and choose some integer upper bounds. Therefore, we have

$$\delta = \delta_1 + \delta_2 = \frac{2\sum\limits_{i=1}^{\alpha+\beta} |\mathcal{R}_i|}{N} + \frac{2\sum_{i=1}^{\alpha} c_i^2}{N} + \frac{2q_c^2(\sum_{i=1}^{\alpha} c_i^2) + 4q_c q^2}{N^2} + \frac{20q^4}{N^3}.$$

This concludes the proof. □

4.4 Proof of Mirror Theory - Upper Bound for $\xi_{\max} = 2$

Theorem 8 (Upper Bound Mirror Theory). *Let $\mathcal{G}$ be a nice graph and q denote the number of edges of $\mathcal{G}$ for $q \le \frac{N}{13}$. It holds that*

$$\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} \le \exp\left(\frac{3\sum_{i=1}^{q}|\mathcal{R}_i|}{N} + \frac{72q^3}{N^3} + \frac{10(n+1)^2}{N}\right).$$

To prove this theorem, we first state the following lemma:

Lemma 20. *For $i \in [2n+2, \beta-1]$, it holds that*

$$h_{i+1} \le (N - 2C_i + |\mathcal{R}_{i+1}|(1 + \frac{116q^2}{N^2}) + \frac{C_i^2}{N} + \frac{124q^2}{N^2})h_i$$

Proof. For $V, V' \in \mathbb{L}_{i+1}$, if $i \ge 2n+2$, by Lemma 17, then we have

$$\left|\frac{h_i}{N} - h'(V,V')\right| \le \frac{(29\,|\mathcal{R}_{i+1}| + 31)h_i}{N^2},$$

equivalently, we have

$$h'(V,V') \le \frac{h_i}{N}\left(1 + \frac{29\,|\mathcal{R}_{i+1}| + 31}{N}\right)$$

Plugging in Equation (24), we have

$$\begin{aligned}\sum_{(V_1,\dots,V_i)\in\mathcal{S}_i} |\varLambda_i \bigcap (\varLambda_i \oplus \lambda_i)| &= |\mathcal{R}_{i+1}|h_i + \sum_{\{V,V'\}\in\mathbb{L}_{i+1}} h'(V,V') \\ &\le \left(|\mathcal{R}_{i+1}| + \frac{C_i^2}{N}\left(1 + \frac{29\,|\mathcal{R}_{i+1}| + 31}{N}\right)\right) h_i.\end{aligned}$$

Plugging the above inequality into Equation (23), we have

$$h_{i+1} \le \left(N - 2C_i + |\mathcal{R}_{i+1}|\left(1 + \frac{116q^2}{N^2}\right) + \frac{C_i^2}{N} + \frac{124q^2}{N^2}\right) h_i.$$

This completes the proof. $\square$

Using Lemma 20, we can prove Theorem 8 as follows.

Proof (of Theorem 8). Recall when $q_c = 0$, $\alpha = 0$. To recursively compute the upper bound for $\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_\beta}}$, we first upper bound $\frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)}$ for $i = 2n+2, \cdots, q$. For $i = 2n+2, \cdots, q$, using Lemma 20, we have

$$\begin{aligned}\frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)} &\le \frac{N^2 - 2C_iN + |\mathcal{R}_{i+1}|(N + \frac{116q^2}{N}) + \frac{C_i^2}{2} + \frac{124q^2}{N}}{N^2 - (2C_i - 1)N + C_i^2 + C_i} \\ &\le 1 + \frac{|\mathcal{R}_{i+1}|(N + \frac{116q^2}{N}) + \frac{124q^2}{N}}{N^2 - (2C_i-1)N + C_i^2 + C_i} \\ &\le 1 + \frac{2|\mathcal{R}_{i+1}|}{N} + \frac{138|\mathcal{R}_{i+1}|q^2}{N^3} + \frac{147q^2}{N^3} \\ &\le 1 + \frac{3|\mathcal{R}_{i+1}|}{N} + \frac{147q^2}{N^3}.\end{aligned}$$

By using the above inequality, we have

$$\begin{aligned}
\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_\beta}} &= \prod_{i=0}^{2n+1} \frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)} \times \prod_{i=2n+2}^{q-1} \frac{h_{i+1}(N-1)}{h_i(N-C_i)(N-C_i-1)} \\
&\leq \prod_{i=0}^{2n+1} \left(1 + \frac{2C_iN}{N^2-(2C_i-1)N+C_i^2+C_i}\right) \\
&\quad \times \prod_{i=2n+2}^{q-1} \left(1 + \frac{3|\mathcal{R}_{i+1}|}{N} + \frac{147q^2}{N^3}\right) \qquad (\because h_{i+1} \leq Nh_i) \\
&\leq \prod_{i=0}^{2n+1} \left(1+\frac{5i}{N}\right) \times \left(1 + \frac{3\sum_{i=1}^{q}|\mathcal{R}_i|}{Nq} + \frac{147q^2}{N^3}\right)^q \\
&\qquad (\because C_i \leq \min\{i, N/13\} \text{ and by Jensen's Inequality}) \\
&\leq \left(1 + \frac{10(n+1)^2}{N}\right) e^{\delta_1} \qquad \left(\text{substitute } \delta_1 = \tfrac{3\sum_{i=1}^{q}|\mathcal{R}_i|}{N} + \tfrac{147q^3}{N^3}\right) \\
&\leq e^{\delta_2+\delta_1}, \qquad \left(\text{substitute } \delta_2 = \tfrac{10(n+1)^2}{N}\right)
\end{aligned}$$

which completes the proof. □

5 Multi-User Security of XoP

This section proves the multi-user PRF* security of XoP such that

$$\mathsf{XoP}[\mathsf{P}](x) := P(0\|x) \oplus P(1\|x)$$

where P is an n-bit random permutation.

The first theorem follows the paradigm of [13], the Chi-squared method. We slightly adapted the proof because we are not considering the truncation. This is relatively weaker than the result obtained from the Squared-ratio method described below, but effectively exemplifies the power of our new ideal world without outputting zero.

Theorem 9. *Let n, u, and q_m be positive integers such that $q_m \leq 2^{n-3}$. Then, it holds*

$$\mathsf{Adv}_{\mathsf{XoP}}^{\mathsf{mu\text{-}prf}^*}(u, q_m) \leq \left(\frac{6uq_m^3}{(2^n-1)^3}\right)^{\frac{1}{2}}.$$

The second result is obtained by following the paradigm of [11]. We stress that we have no intention to optimize the constant factors, and there is significant room for improvement on them.

Theorem 10. *Let n, u, and q_m be positive integers such that $n > 12$ and $q_m \leq \frac{2^n}{4n}$. Then, it holds*

$$\mathsf{Adv}_{\mathsf{XoP}}^{\mathsf{mu\text{-}prf}^*}(u, q_m) \leq \frac{26u^{\frac{1}{2}}q_m^2}{2^{2n}} + \frac{49u^{\frac{1}{2}}(n+1)^2}{2^n}.$$

The remainder of this section is organized as follows. We first prove Theorem 9 in Section 5.1 using the Chi-squared method. In Section 5.2, we prove Theorem 10 using the Squared-ratio method.

5.1 Proof of Multi-User Security of XoP via the Chi-Squared Method

Before proving the security of XoP, we define multiple experiments. First, let $\mathcal{S}_0$ be the fine-tuned ideal world where each user interacts with the random function sampled from $\mathsf{Func}^*(n-1,n)$. The world $\mathcal{S}_1$ is defined similarly, but each user interacts with different XoP constructions. An adversary makes q_m queries to each user interface, a total of $q = uq_m$ queries. Without loss of generality, we assume an information-theoretic adversary is deterministic and does not make any redundant query; any redundant query only degrades the adversary's ability.

We consider the experiments in Algorithm 1 that are essentially identical to the original worlds but with lazily sampling the queries; each answer for the oracle queries of the j-th user is replaced by the output $\mathbf{Z}^j$. This change cannot be observed from the adversary's view. Thus, we have

$$\|\mathsf{p}_{\mathcal{S}_0}(\cdot) - \mathsf{p}_{\mathcal{S}_1}(\cdot)\| = \|\mathsf{p}_{\mathcal{B}_0}(\cdot) - \mathsf{p}_{\mathcal{B}_1}(\cdot)\|.$$

Algorithm 1 Ideal/Real experiments for XoP

Experiment $\mathcal{B}_0$

for $j \leftarrow 1$ to u **do**
  for $i \leftarrow 1$ to q_m **do**
    $y_i^j \leftarrow_\$ \{0,1\}^n \setminus \{\mathbf{0}\}$
  $\mathbf{Z}^j \leftarrow (y_1^j, \ldots, y_{q_m}^j)$
return $(\mathbf{Z}^1, \ldots, \mathbf{Z}^u)$

Experiment $\mathcal{B}_1$

for $j \leftarrow 1$ to u **do**
  $\mathcal{R} \leftarrow \{0,1\}^n$
  for $i \leftarrow 1$ to q_m **do**
    $u_{2i-1}^j \leftarrow_\$ \mathcal{R}$, $\mathcal{R} \leftarrow \mathcal{R} \setminus \{u_{2i-1}^j\}$
    $u_{2i}^j \leftarrow_\$ \mathcal{R}$, $\mathcal{R} \leftarrow \mathcal{R} \setminus \{u_{2i}^j\}$
    $y_i^j \leftarrow u_{2i-1}^j \oplus u_{2i}^j$
  $\mathbf{Z}^j \leftarrow (y_1^j, \ldots, y_{q_m}^j)$
return $(\mathbf{Z}^1, \ldots, \mathbf{Z}^u)$

We then consider the intermediate worlds in Algorithm 2 for applying the Chi-squared method. In the world $\mathcal{C}_0$, the oracle pretends to the answer y is of the form $u \oplus u'$ for $u = P(0\|x)$ and $u' = P(1\|x)$ for the given permutation P

and an input x, and returns $z = (u, u')$. When the adversary processes it as if the oracle outputs $y = u \oplus u'$, and as long as the oracle does not return $(\perp, \perp)$, this world is identical to $\mathcal{B}_0$ in the adversary's view. In the world $\mathcal{C}_1$, everything is the same with $\mathcal{B}_1$, but the oracle returns (u, u') as in $\mathcal{C}_0$. The following lemma

Algorithm 2 Intermediate experiments for XoP

Experiment $\mathcal{C}_0$

for $j \leftarrow 1$ to u **do**
    $\mathcal{R} \leftarrow \{0,1\}^n$
    for $i \leftarrow 1$ to q_m **do**
        $y_i^j \leftarrow_{\$} \{0,1\}^n \setminus \{\mathbf{0}\}$
        $\mathcal{T}_i^j(y_i^j) \leftarrow \{(u,v) : u, v \in \mathcal{R},\ u \neq v,\ u \oplus v = y_i^j\}$
        if $\left|\mathcal{T}_i^j(y_i^j)\right| > 0$ **then**
            $(u_{2i-1}^j, u_{2i}^j) \leftarrow_{\$} \mathcal{T}_i^j(y_i^j)$
        else
            $(u_{2i-1}^j, u_{2i}^j) \leftarrow (\perp, \perp)$
        $\mathcal{R} \leftarrow \mathcal{R} \setminus \{u_{2i-1}^j, u_{2i}^j\}$
        $z_i^j \leftarrow (u_{2i-1}^j, u_{2i}^j)$
    $\mathbf{Z}^j \leftarrow (z_1^j, \ldots, z_{q_m}^j)$
return $(\mathbf{Z}^1, \ldots, \mathbf{Z}^u)$

Experiment $\mathcal{C}_1$

for $j \leftarrow 1$ to u **do**
    $\mathcal{R} \leftarrow \{0,1\}^n$
    for $i \leftarrow 1$ to q_m **do**
        $u_{2i-1}^j \leftarrow_{\$} \mathcal{R}$, $\mathcal{R} \leftarrow \mathcal{R} \setminus \{u_{2i-1}^j\}$
        $u_{2i}^j \leftarrow_{\$} \mathcal{R}$, $\mathcal{R} \leftarrow \mathcal{R} \setminus \{u_{2i}^j\}$
        $z_i^j \leftarrow (u_{2i-1}^j, u_{2i}^j)$
    $\mathbf{Z}^j \leftarrow (z_1^j, \ldots, z_{q_m}^j)$
return $(\mathbf{Z}^1, \ldots, \mathbf{Z}^u)$

holds for Experiment $\mathcal{C}_0$ in Algorithm 2.

Lemma 21. *If $q_m \leq 2^{n-3}$, Experiment $\mathcal{C}_0$ in Algorithm 2 never returns $(\perp, \perp)$.*

Proof. We suppose any $j \in [u]$ and omit y for simplicity. If $i = 1$, it is trivial that $|\mathcal{T}_i(y_i)| = 2^n > 0$. For $2 \leq i \leq q_m$, we have $|\mathcal{R}| = 2^n - 2(i-1)$. Note that $(u, v) \in \mathcal{T}_i(y_i) \Rightarrow (v, u) \in \mathcal{T}_i(y_i)$. Therefore $|\mathcal{T}_i(y_i)| \geq 2^n - 4(i-1) > 2^{n-1} > 0$ since $i \leq q_m \leq 2^{n-3}$. $\square$

As described above, we can regard an adversary in $\mathcal{B}_0$ and $\mathcal{B}_1$ as special cases of $\mathcal{C}_0$ and $\mathcal{C}_1$ if $\mathcal{C}_0$ never returns $(\perp, \perp)$, we have the following inequality:

$$\|\mathsf{p}_{\mathcal{S}_0}(\cdot) - \mathsf{p}_{\mathcal{S}_1}(\cdot)\| = \|\mathsf{p}_{\mathcal{B}_0}(\cdot) - \mathsf{p}_{\mathcal{B}_1}(\cdot)\| \leq \|\mathsf{p}_{\mathcal{C}_0}(\cdot) - \mathsf{p}_{\mathcal{C}_1}(\cdot)\|. \tag{26}$$

Concluding the proof with the Chi-squared method. Without loss of the generality, assume that each user makes q_m queries. For $i \in [q]$ where $i = (j-1)q_m + k$ such that $j \in [u]$ and $k \in [q_m]$, the response of the i-th query is seen as $z_i = z_k^j$. We can easily check that the support of $\mathsf{p}_{\mathcal{C}_1}^{i-1}(\cdot)$ is contained in the support of $\mathsf{p}_{\mathcal{C}_0}^{i-1}(\cdot)$ for $i = 1, \ldots, q$, allowing us to use the Chi-squared method. Let $\Omega = \{0,1\}^n \times \{0,1\}^n$. For a fixed $i \in \{1, \ldots, q\}$, let $i \in [q]$ where $i = (j-1)q_m + k$ such that $j \in [u]$ and $k \in [q_m]$. Fix $\mathbf{z} \in \Omega^{i-1}$ such that $\mathsf{p}_{\mathcal{C}_1}^{i-1}(\mathbf{z}) > 0$. We will compute

$$\chi^2(\mathbf{z}) = \sum_{\substack{z \in \Omega \text{ such that} \\ \mathsf{p}_{\mathcal{C}_0,i}^{\mathbf{z}}(z) > 0}} \frac{\left(\mathsf{p}_{\mathcal{C}_1,i}^{\mathbf{z}}(z) - \mathsf{p}_{\mathcal{C}_0,i}^{\mathbf{z}}(z)\right)^2}{\mathsf{p}_{\mathcal{C}_0,i}^{\mathbf{z}}(z)}.$$

Note that z_i is independent with $\mathbf{Z}^{(j-1)q_m}$. For $y \in \{0,1\}^n$, let $T_k^j(y) = \left|\mathcal{T}_k^j(y)\right|$. Also note that $\mathcal{T}_k^j(y)$ can be defined in both $\mathcal{C}_0$ and $\mathcal{C}_1$. From the proof of Lemma 21, we have $T_k^j(y) \geq 2^n - 4(k-1)$. Moreover, we see that

$$\begin{aligned}
\mathsf{p}_{\mathcal{C}_0,i}^{\mathbf{z}}(u,v) &= \frac{1}{(2^n-1)T_k^j(y)}, \\
\mathsf{p}_{\mathcal{C}_1,i}^{\mathbf{z}}(u,v) &= \frac{1}{(2^n-2k+2)(2^n-2k+1)} = \frac{1}{(2^n-2k+2)_2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\chi^2(\mathbf{z}) &= \sum_{\substack{(u,v) \in \Omega \text{ such} \\ \text{that } \mathsf{p}_{\mathcal{C}_0,i}^{\mathbf{z}}(u,v) > 0}} \frac{(2^n-1)\left(T_k^j(y) - \frac{(2^n-2k+2)_2}{2^n-1}\right)^2}{T_k^j(y)(2^n-2k+2)^2(2^n-2k+1)^2} \\
&\leq \frac{32}{9} \cdot \frac{2^n-1}{2^{3n}(2^n-2k+2)_2} \sum_{\substack{(u,v) \in \Omega \text{ such} \\ \text{that } \mathsf{p}_{\mathcal{C}_0,i}^{\mathbf{z}}(u,v) > 0}} \left(T_k^j(y) - \frac{(2^n-2k+2)_2}{2^n-1}\right)^2 \\
&= \frac{32}{9} \cdot \frac{2^n-1}{2^{2n}(2^n-2k+2)_2} \sum_{y \in \{0,1\}^n} \left(T_k^j(y) - \frac{(2^n-2k+2)_2}{2^n-1}\right)^2. \qquad (27)
\end{aligned}$$

since $k \leq q_m \leq 2^{n-3}$. We claim the following lemma proved in Section 5.1.1.

Lemma 22. *It holds that*

$$\begin{aligned}
\mathbf{Ex}\left[T_k^j(y)\right] &= \frac{(2^n-2k+2)_2}{2^n-1}, \\
\mathbf{Var}\left[T_k^j(y)\right] &= \frac{2(2k-2)(2k-3)(2^n-2k+2)_2}{(2^n-1)^2(2^n-3)}
\end{aligned}$$

where the expectation and variance are taken over from the distribution of $\mathcal{C}_1$.

From Equation (27) and Lemma 22, it follows that

$$\begin{aligned}
\mathbf{Ex}\left[\chi^2(\mathbf{z})\right] &\le \frac{32}{9}\cdot\frac{2^n-1}{2^{2n}(2^n-2k+2)_2}\mathbf{Ex}\left[\sum_{y\in\{0,1\}^n}\left(T_k^j(y)-\frac{(2^n-2k+2)_2}{2^n-1}\right)^2\right]\\
&=\frac{32}{9}\cdot\frac{2^n-1}{2^{n}(2^n-2k+2)_2}\mathbf{Var}\left[T_k^j(y)\right]\\
&=\frac{32}{9}\cdot\frac{2(2k-2)(2k-3)}{2^n(2^n-1)(2^n-3)}\le\frac{32(k-1)^2}{(2^n-1)^3}
\end{aligned}$$

and finally, we have the following inequality, which concludes the proof:

$$\begin{aligned}
\|\mathsf{p}_{\mathcal{C}_0}(\cdot)-\mathsf{p}_{\mathcal{C}_1}(\cdot)\| &\le \left(\frac{1}{2}\sum_{i=1}^{q}\mathbf{Ex}\left[\chi^2(\mathbf{z})\right]\right)^{\frac{1}{2}} = \left(\frac{1}{2}\sum_{j=1}^{u}\sum_{k=1}^{q_m}\mathbf{Ex}\left[\chi^2(\mathbf{z})\right]\right)^{\frac{1}{2}}\\
&\le\left(\frac{1}{2}\sum_{j=1}^{u}\sum_{k=1}^{q_m}\frac{32(k-1)^2}{(2^n-1)^3}\right)^{\frac{1}{2}}\le\left(\frac{6uq_m^3}{(2^n-1)^3}\right)^{\frac{1}{2}}.
\end{aligned}$$

5.1.1 Proof of Lemma 22 Let $\Psi=\{0,1\}^n$ and fix j and k. Let I_ψ where $\psi\in\Psi$ be an indicator variable

$$I_\psi=1\Leftrightarrow\psi,\psi\oplus y\in\{0,1\}^n\setminus\{u_\ell^j\}_{\ell\in[2k-2]}.$$

Note that the variables $\{u_j^\ell\}$ are uniformly randomly sampled from $\{0,1\}^n$ without replacement. Observe that

$$T_k^j(y)=\sum_{\psi\in\Psi}I_\psi$$

and

$$\mathbf{Ex}\left[I_\psi\right]=\frac{(2^n-2k+2)(2^n-2k+1)}{2^n(2^n-1)}.$$

Thus, we have

$$\mathbf{Ex}\left[T_k^j(y)\right]=\sum_{\psi\in\Psi}\mathbf{Ex}\left[I_\psi\right]=\frac{(2^n-2k+2)(2^n-2k+1)}{2^n-1}. \tag{28}$$

Now, we compute the following expectation

$$\mathbf{Ex}\left[\left(T_k^j(y)\right)^2\right]=\mathbf{Ex}\left[\left(\sum_{\psi\in\Psi}I_\psi\right)^2\right]=\mathbf{Ex}\left[\sum_{(\psi,\psi')\in\Psi^2}I_\psi I_{\psi'}\right].$$

For ψ and ψ', let r be the size of the following set

$$\{\psi, \psi', \psi \oplus y, \psi' \oplus y\}.$$

We see that $r \in \{2, 4\}$ since $y \neq \mathbf{0}$; moreover,

$$\mathbf{Ex}\left[I_\psi I_{\psi'}\right] = \frac{(2^n - 2k + 2)_r}{(2^n)_r}.$$

For a fixed $\psi \in \Psi$, we have

$$|\{\psi' \in \Psi \mid r = 2\}| = 2,$$
$$|\{\psi' \in \Psi \mid r = 4\}| = 2^n - 2.$$

It follows that

$$\sum_{\substack{\psi' \in \Psi, \\ r=2}} \mathbf{Ex}\left[I_\psi I_{\psi'}\right] = \frac{2(2^n - 2k + 2)_2}{(2^n)_2},$$

$$\sum_{\substack{\psi' \in \Psi, \\ r=4}} \mathbf{Ex}\left[I_\psi I_{\psi'}\right] = (2^n - 2)\left(1 - \frac{2k-2}{2^n - 2}\right)\left(1 - \frac{2k-2}{2^n - 3}\right)\frac{(2^n - 2k + 2)_2}{(2^n)_2}.$$

As $\mathbf{Ex}\left[\sum_{(\psi,\psi') \in \Psi^2} I_\psi I_{\psi'}\right] = \sum_{(\psi,\psi') \in \Psi^2} \mathbf{Ex}\left[I_\psi I_{\psi'}\right] = \sum_{\psi \in \Psi} \sum_{\psi' \in \Psi} \mathbf{Ex}\left[I_\psi I_{\psi'}\right]$ and the sum is divided into two cases according to the value of r, the sum of the expectations is given as

$$\mathbf{Ex}\left[\sum_{(\psi,\psi') \in \Psi^2} I_\psi I_{\psi'}\right] = 2^n \left(\sum_{\substack{\psi' \in \Psi, \\ r=2}} \mathbf{Ex}\left[I_\alpha I_{\psi'}\right] + \sum_{\substack{\psi' \in \Psi, \\ r=4}} \mathbf{Ex}\left[I_\alpha I_{\psi'}\right]\right). \tag{29}$$

for an arbitrary constant $\alpha \in \Psi$. Therefore, by Equation (29), we have

$$\begin{aligned}
\mathbf{Ex}\left[\sum_{(\psi,\psi') \in \Psi^2} I_\psi I_{\psi'}\right] &= \frac{(2^n - 2k + 2)_2}{2^n - 1}\left(2^n - (2k-2)\left(2 + \frac{1}{2^n - 3}\right) + \frac{(2k-2)^2}{2^n - 3}\right) \\
&= \frac{(2^n - 2k + 2)_2}{2^n - 1}\left(2^n - 4k + 4 + \frac{(2k-2)_2}{2^n - 3}\right).
\end{aligned}$$

for a fixed ψ. By Equation (28), conclude that

$$\begin{aligned}
\mathbf{Var}\left[T_k^j(y)\right] &= \mathbf{Ex}\left[\sum_{(\psi,\psi') \in \Psi^2} I_\psi I_{\psi'}\right] - \left(\mathbf{Ex}\left[\sum_{\psi \in \Psi} I_\psi\right]\right)^2 \\
&= \frac{(2^n - 2k + 2)_2}{2^n - 1}\left(2^n - (4k-4) + \frac{(2k-2)_2}{2^n - 3} - \frac{(2^n - 2k + 2)_2}{2^n - 1}\right) \\
&= \frac{(2^n - 2k + 2)_2}{2^n - 1} \cdot \frac{2(2k-2)(2k-3)}{(2^n - 1)(2^n - 3)} \\
&= \frac{2(2k-2)(2k-3)(2^n - 2k + 2)_2}{(2^n - 1)^2(2^n - 3)}
\end{aligned} \tag{30}$$

By Equations (28) and (30), the proof completes.

5.2 Proof of Multi-User Security of XoP via the Squared-Ratio Method

Thanks to the squared-ratio method (Theorem 2), considering an adversary $\mathcal{D}$ making at most q_m queries in the information-theoretic setting suffices. The queries made by $\mathcal{D}$ are $x_1, ..., x_{q_m} \in \{0,1\}^{n-1}$, which are assumed to be all different without loss of the generality. In this way, $\mathcal{D}$ obtains a transcript $\tau = ((x_1, \lambda_i), \cdots, (x_{q_m}, \lambda_{q_m}))$.

In the real world, $\mathsf{XoP}[\mathsf{P}](x_i) \overset{\text{def}}{=} P_{\gamma_i} \oplus P_{\gamma_i'}$, where P is a given n-bit (keyed) PRP function, and $\{P_{\gamma_1}, P_{\gamma_1'}, \cdots, P_{\gamma_n}, P_{\gamma_n'}\}$ should be a solution to the following equation system

$$\Gamma =: \begin{cases} P_{\gamma_1} \oplus P_{\gamma_1'} = \lambda_1, \\ P_{\gamma_2} \oplus P_{\gamma_2'} = \lambda_2, \\ \qquad \vdots \\ P_{\gamma_{q_m}} \oplus P_{\gamma_{q_m}'} = \lambda_{q_m}. \end{cases}$$

This induces the transcript graph $\mathcal{G}(\tau)$ in the real world.

Bad transcript analysis. Recall Theorem 4. To upper bound $|\mathcal{R}_i|$ corresponding to this system in the ideal world, where each λ_j takes an independent random value sampled from $\{0,1\}^n \setminus \{\mathbf{0}\}$, we define a bad event as follows:

bad: $\exists (i_1, \cdots, i_n) \in [q_m]^{*n}$ such that $\lambda_{i_1} = \cdots = \lambda_{i_n}$.

We have

$$\Pr[\mathsf{bad}] = \frac{\binom{q_m}{n}}{(2^n-1)^{n-1}} \leq \frac{q_m^n}{n!(2^n-1)^{n-1}} \leq \left(\frac{q_m}{2^n}\right)^n.$$

because $n! \geq 2^{n+1}$ and $2 \cdot (2^n-1)^{n-1} \geq (2^n)^n$ for $n > 12$. We say that the transcript is good if it is not bad.

Good transcript analysis. Now we focus on the good transcript. Let T_{id} and T_{re} be random variables following the distribution of the transcripts in the real world and the ideal world, respectively. Then, we have

$$\frac{\Pr[\mathsf{T}_{\mathsf{re}} = \tau]}{\Pr[\mathsf{T}_{\mathsf{id}} = \tau]} = \frac{h(\mathcal{G})(2^n-1)^{q_m}}{(2^n)_{2q_m}},$$

because in the real world, the transcript can occur with a probability proportional to the number of solutions $h(\mathcal{G})$ over the choices of $P_{\gamma_i}, P_{\gamma_i'}$. On the other hand, in the ideal world, every transcript occurs equally, i.e., with probability $\frac{1}{(2^n-1)^{q_m}}$.

Recall the following definition of $\mathcal{R}_i$, where $|\mathcal{R}_i| \leq n$ for all $i \in [q_m]$ because of $\neg\mathsf{bad}$ in our case:

$$\mathcal{R}_i = \left\{(V_1, V_1', V_2, V_2') \in {\mathcal{C}_i}^{*2} \times {\mathcal{C}_j}^{*2} \,\middle|\, j < i \text{ and } \lambda(V_1, V_1') = \lambda(V_2, V_2')\right\}.$$

Let $I_{j,k}$ be a random variable that equals 1 if $\lambda_i = \lambda_j$ and 0 otherwise. Then it holds that $\sum_{i=1}^{q_m} |\mathcal{R}_i| = \sum_{j,k} I_{j,k}$ as a random variable, which will be used later. The bound of $\neg\mathsf{bad}$ implies that

$$\frac{3\sum_{i=1}^{q_m} |\mathcal{R}_i|}{2^n} + \frac{3q_m^2}{2^{2n}} + \frac{10(n+1)^2}{2^n} \leq \frac{3nq_m}{2^n} + \frac{3q_m^2}{2^{2n}} + \frac{10(n+1)^2}{2^n} \leq 1$$

for $n > 12$ and $q_m \leq \frac{2^n}{4n}$. Therefore, by Theorem 4,

$$\left| \frac{h(\mathcal{G})(2^n-1)^{q_m}}{(2^n)_{2q_m}} - 1 \right| \leq \frac{6\sum_{i=1}^{q_m} |\mathcal{R}_i|}{2^n} + \frac{6{q_m}^2}{2^{2n}} + \frac{20(n+1)^2}{2^n}.$$

Conclude the proof. Define $\epsilon_2 = \left(\frac{q_m}{2^n}\right)^n$ and

$$\epsilon_1(\tau) = \frac{6\sum_{i=1}^{q_m} |\mathcal{R}_i|}{2^n} + \frac{6{q_m}^2}{2^{2n}} + \frac{20(n+1)^2}{2^n}.$$

To apply Theorem 2, we need to bound the expectation of $\epsilon_1(\tau)^2$ where the randomness is taken over the distribution of the ideal world. We apply Lemma 6, and using Lemma 5 for bounding the expectations relevant to $\mathcal{R}_i$, we have

$$\begin{aligned}
\mathbf{Ex}\left[\epsilon_1(\tau)^2\right] &\leq \frac{108\mathbf{Ex}\left[\left(\sum_{i=1}^{q_m} |\mathcal{R}_i|\right)^2\right]}{2^{2n}} + \frac{108q_m^4}{2^{4n}} + \frac{1200(n+1)^4}{2^{2n}} \\
&\leq \frac{108q_m^2}{2^{2n}(2^n-1)} + \frac{108q_m^4}{2^{2n}(2^n-1)^2} + \frac{108q_m^4}{2^{4n}} + \frac{1200(n+1)^4}{2^{2n}} \\
&\leq \frac{318q_m^4}{2^{4n}} + \frac{1200(n+1)^4}{2^{2n}}.
\end{aligned}$$

Applying Theorem 2 and plugging in the above inequality, we have

$$\mathsf{Adv}_{\mathsf{XoP}}^{\mathsf{mu\text{-}prf}^*}(u, q_m) \leq \sqrt{2u\mathbf{Ex}\left[\epsilon_1^2\right]} + 2u\epsilon_2 \leq \frac{26u^{\frac{1}{2}}q_m^2}{2^{2n}} + \frac{49u^{\frac{1}{2}}(n+1)^2}{2^n}.$$

This completes the proof. □

6 Multi-User Security of nEHtM

This section proves the multi-user MAC security of the nonce-based Enhanced Hash-then-mask (nEHtM) scheme proposed by [23]. Let H be a $(n-1)$-bit output δ-AXU hash function and let P be an n-bit permutation. For given inputs a message M and an $(n-1)$-bit nonce N, $\mathsf{nEHtM} = \mathsf{nEHtM}[\mathsf{H}, \mathsf{P}]$ outputs a tag T defined as follows:

$$T = \mathsf{nEHtM}(N, M) := \mathsf{P}(0\|N) \oplus \mathsf{P}(1\|\mathsf{H}_{K_h}(M) \oplus N).$$

An adversary $\mathcal{A}$ for the nEHtM makes two types of queries: MAC queries that compute the tags given inputs messages and nonces, and verification queries that take a tuple of a nonce, a message, and a candidate tag (N', M', T') as inputs and is returned $b \in \{0,1\}$, where $b = 1$ if and only if the equation $\mathsf{nEHtM}(N', M') = T'$ holds. The main result of this section is summarized as follows.

Theorem 11. *Let $n \geq 20$ be a positive integer. Let $\delta > 0$ and $\mathsf{H} : \mathcal{K} \times \mathcal{M} \to \{0,1\}^{n-1}$ be a δ-AXU hash function family. Let u, q_m, v_m, and μ_m be positive integers such that $uq_m^2\delta^2 \leq 1$ and $32\mu_m q_m \leq 1$. Then, $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{nEHtM}}(u, \mu_m, q_m, v_m)$ is bounded by*

$$72u\mu_m^2\delta + \frac{140n\sqrt{u}\mu_m^2}{2^n} + 129uv_m\delta + 149 \cdot \left(\frac{uq_m^4\delta}{2^{2n}}\right)^{\frac{1}{2}} + 80 \cdot \left(\frac{n^2u\mu_m^2q_m^3\delta}{2^{2n}}\right)^{\frac{1}{2}}$$
$$+ 12 \cdot \left(\frac{u^2\mu_m q_m^3}{2^{3n}}\right)^{\frac{1}{2}} + 153 \cdot \left(\frac{n^2u^2\mu_m^2q_m^2\delta}{2^{2n}}\right)^{\frac{1}{3}} + 155 \cdot \left(\frac{n^2u^2q_m^5\delta^2}{2^{2n}}\right)^{\frac{1}{3}}$$

Assuming $\delta = \frac{\ell}{2^n}$ for some $\ell \geq 1$, we have the following asymptotic bound:

$$O\left(\frac{\ell u(n\mu_m^2 + v_m)}{2^n} + \frac{\ell^{\frac{1}{2}}nu\mu_m q_m^{\frac{3}{2}}}{2^{\frac{3n}{2}}} + \frac{\ell^{\frac{1}{2}}u^{\frac{1}{2}}q_m^2}{2^{\frac{3n}{2}}} + \left(\frac{\ell n^2u^2\mu_m^2q_m^2}{2^{3n}}\right)^{\frac{1}{3}} + \left(\frac{\ell^2n^2u^2q_m^5}{2^{4n}}\right)^{\frac{1}{3}}\right).$$

Plugging $\mu_m = 0, v_m = 0$ in this bound matches the nonce-respecting security bound, resulting in the asymptotic bound

$$\tilde{O}\left(\left(\frac{uq_m^4}{2^{3n}}\right)^{\frac{1}{2}} + \left(\frac{u^2q_m^5}{2^{4n}}\right)^{\frac{1}{3}}\right),$$

ignoring small factors, which is more carefully dealt in Section 6.4. Figure 2 shows the graphical comparison between our bounds and the previous bounds [11, 16] in this setting.

We further explore the multi-user security of nEHtM with hash functions with a stronger property, dubbed a pairwise δ-almost XOR universal: for any $M_1 \neq M_1'$ and $M_2 \neq M_2'$ in $\mathcal{M}$ such that $\{M_1, M_1'\} \neq \{M_2, M_2'\}$ and $X_1, X_2 \in \mathcal{X}$, it holds that

$$\Pr_{K \xleftarrow{\$} \mathcal{K}}\left[\mathsf{H}_{K_h}(M_1) \oplus \mathsf{H}_{K_h}(M_1') = X_1 \wedge \mathsf{H}_{K_h}(M_2) \oplus \mathsf{H}_{K_h}(M_2') = X_2\right] \leq \delta^2.$$

In this setting, we obtain a much better bound on $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{nEHtM}}(u, \mu_m, q_m, v_m)$ of

$$\tilde{O}\left(\frac{u\mu_m^2 + uv_m}{2^n} + \frac{\sqrt{u}q_m^4}{2^{3n}} + \left(\frac{u^2\mu_m^2q_m^2}{2^{3n}}\right)^{1/3} + \left(\frac{uq_m^2}{2^{2n}}\right)^{2/3} + \left(\frac{u^2q_m^6}{2^{5n}}\right)^{1/3}\right)$$

ignoring polynomial factors of ℓ and n for $\delta = O(\ell/2^n)$. For the mu PRF security, we have the following security bound assuming the strong hash functions:

$$\mathsf{Adv}^{\mathsf{mu\text{-}prf}^*}_{\mathsf{nEHtM}}(u, q_m) = \tilde{O}\left(\frac{\sqrt{u}q_m^4}{2^{3n}} + \left(\frac{uq_m^2}{2^{2n}}\right)^{2/3} + \left(\frac{u^2q_m^6}{2^{5n}}\right)^{1/3}\right).$$

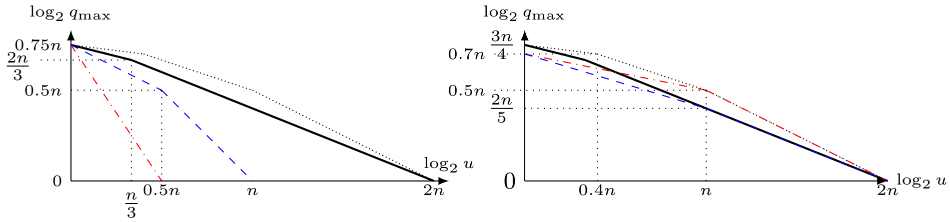

Fig. 2: Comparison of the multi-user security bounds (in terms of the threshold number of queries per user) as functions of $\log_2 u$. We neglect the polynomial terms of ℓ and $\log n$ in the graphs. We assume $v_m = \mu_m = 0$ for a fair comparison. The solid line represents our bounds in both graphs. In the left graph, the blue dashed line (resp. the red dash-dotted line) represents the security bound obtained by the hybrid argument where $q = q_m$ (resp. $q = uq_m$). On the other hand, in the right graph, the blue dashed line corresponds to the result of [11] with our correction in Section 6.5. The red dash-dotted line in the right graph corresponds to the claimed security bound in [11], which was buggy. Assuming δ-AXU$^{(2)}$, the dash-dotted line is recovered, while the densely dotted line can be proven with the method in this paper.

In the remainder of this section, we prove Theorem 11 using Theorems 2 and 3. The proof sketch of the nonce-respecting setting can be found in Section 6.4. The stronger bound with a pairwise δ-AXU and the discussion on the previous $\mathsf{nEHtM2}$ security proof [11] is placed in Section 6.5.

Before starting the proof, some observations are in order. First, we always assume that $q_m \le \frac{2^{3n/4}}{8} \le \frac{2^n}{256}$, $\mu_m \le \frac{2^{0.5n}}{12\sqrt{n}}, \frac{2^n}{32q_m}$, $nuq_m\delta < 2^n$, and $v_m \le \frac{2^n}{128}$, otherwise the right hand side of the advantage becomes ≥ 1, and nothing to prove. Second, we do not intend to optimize the constant factors in the proof and sometimes even give up on optimizing the small factors ℓ and n. The constants between the inequalities may be chosen as a rough upper bound.

6.1 Bad and Good Transcripts

The queries of the adversary can be represented by the MAC queries and the verification queries as follows

$$\tau_m = (N_i, M_i, T_i)_{1\le i\le q_m}, \quad \tau_v = \left(N'_j, M'_j, T'_j, b'_j\right)_{1\le j\le v_m}$$

where $T_i = \mathsf{nEHtM}(N_i, M_i)$ and $b'_j = 1$ if and only if $T'_j = \mathsf{nEHtM}(N'_j, M'_j)$. The overall transcript is

$$\tau = (\tau_m, \tau_v, K)$$

where we assume that the key K is given at the end of the attack for free, which only makes the adversary stronger. We additionally define

$$X_i := \mathsf{H}_{K_h}(M_i) \oplus N_i, \quad X'_j := \mathsf{H}_{K_h}(M'_j) \oplus N'_j$$

for $i = 1, ..., q_m$ and $j = 1, ..., v_m$.

In the real world, these values should obey the following system of equations when the adversary fails to forge the MAC:

$$\begin{cases} \mathsf{P}(0\|N_1) \oplus \mathsf{P}(1\|X_1) = T_1, \\ \mathsf{P}(0\|N_2) \oplus \mathsf{P}(1\|X_2) = T_2, \\ \qquad\vdots \\ \mathsf{P}(0\|N_{q_m}) \oplus \mathsf{P}(1\|X_{q_m}) = T_{q_m}, \end{cases} \quad \text{and} \quad \begin{cases} \mathsf{P}(0\|N'_1) \oplus \mathsf{P}(1\|X'_1) \neq T'_1, \\ \mathsf{P}(0\|N'_2) \oplus \mathsf{P}(1\|X'_2) \neq T'_2, \\ \qquad\vdots \\ \mathsf{P}(0\|N'_{v_m}) \oplus \mathsf{P}(1\|X'_{v_m}) \neq T'_{v_m}. \end{cases}$$

We identify $\{\mathsf{P}(0\|N_i)\}_i \cup \{\mathsf{P}(0\|N'_j)\}_j$ with a set of unknowns

$$\mathcal{P} = \{\mathsf{P}_1, ..., \mathsf{P}_{q_1}\}$$

for $q_1 \leq q_m + v_m$ and similarly identify $\{\mathsf{P}(1\|X_i)\}_i \cup \{\mathsf{P}(1\|X'_j)\}_j$ with a set of unknowns

$$\mathcal{Q} = \{\mathsf{Q}_1, ..., \mathsf{Q}_{q_2}\}$$

for some $q_2 \leq q_m + v_m$.

We define the corresponding transcript graph $\mathcal{G}(\tau) = (\mathcal{V}, \mathcal{E})$ for $\mathcal{V} = \mathcal{P} \sqcup \mathcal{Q}$. Here the set $\mathcal{E}$ includes the following edges: For $i = 1, ..., q_m$, $\mathsf{P}(0\|N_i) \in \mathcal{P}$ and $\mathsf{P}(1\|X_i) \in \mathcal{Q}$ are connected with a $(T_i, =)$-labeled edge. Similarly, for $i = 1, ..., v_m$, $\mathsf{P}(0\|N'_i) \in \mathcal{P}$ and $\mathsf{P}(1\|X'_i) \in \mathcal{Q}$ are connected with $(T'_i, \neq)$-labeled edge. Therefore, the transcript graph $\mathcal{G}(\tau)$ is a connected bipartite graph with two independent sets $\mathcal{P}$ and $\mathcal{Q}$.

In the ideal world, the tags T_i should be a uniformly random element in $\{0,1\}^n \backslash \{\mathbf{0}\}$ and independent from each other; we again stress that the punctured point $\mathbf{0}$ is important of our argument. On the other hand, the candidate tags T'_j are arbitrarily chosen by the adversary from $\{0,1\}^n \backslash \{\mathbf{0}\}$ even in the ideal world.[4] We will compare the difference between the real and ideal worlds regarding the transcript graph $\mathcal{G}(\tau)$.

Notations. Fix a transcript τ so that each N_i, X_i is determined. In the graph $\mathcal{G}^{=}(\tau)$, for each $(n-1)$-bit string $X \in \{0,1\}^{n-1}$, we define the degree of X, denoted by d_X, by the number of $i \in [q_m]$ such that $X_i = X$. We call $(i_1, i_2, ...) \in [q_m]^{*j}$ for some j by a length-j X-trail, which means that it starts from a vertex corresponding to X (see Equation (31)), if

$$(N_{i_1} = N_{i_2}) \wedge (X_{i_2} = X_{i_3}) \wedge ...$$

holds. An X-trail can be interpreted as a trail of

$$\mathsf{P}(1\|X_{i_1}) - \mathsf{P}(0\|N_{i_1}) = \mathsf{P}(0\|N_{i_2}) - ..., \text{ or } X_{i_1} - N_{i_1} = N_{i_2} - ... \tag{31}$$

and similarly define N-trails. (A trail can be both X- and N-trail.) We ambiguously call them trails. Note that a trail (i, j) satisfies $N_i = N_j$ or $X_i = X_j$, and

[4] The adversary can choose $T'_j = \mathbf{0}$. However, the verification query always rejects such a choice, so we ignore this case.

is of length-2. For a trail $\gamma = (i_1, ..., i_j)$, the label of γ is defined by

$$\lambda(\gamma) = \bigoplus_{k \in [j]} T_{i_k},$$

which is equal to $\lambda(V_0, V_\ell)$ for the first and last vertices of γ in the mirror theory. If $\lambda(\gamma) = \lambda(\gamma')$, we say that two trails γ, γ' are a collision pair. A set of trails $\{\gamma_1, ..., \gamma_k\}$ is called by a k-collision if all $\lambda(\gamma_i)$ are equal for all $i \in [k]$. If $\lambda(\gamma) = 0$, γ is called by a null trail.

Bad Transcripts. We first define bad transcripts. Let $L_1, L_2 \geq 2$ be fixed positive integers. Recall q_c denotes the number of edges included in the components of size ≥ 3, and d_X for $X \in \{0,1\}^{n-1}$ denotes the number of $i \in [q_m]$ such that $X_i = X$. We say that the transcript τ is *bad* if any of the following conditions holds. We will choose constants so that $L_1, L_2 \leq \min\left(\frac{2^n}{32 q_m}, \frac{2^{0.5n}}{24\sqrt{n}}\right)$.

- bad_1: $\exists (i,j) \in [q_m]^{*2}$ such that for some $k, \ell \in [q_m]^2$ with $k \neq i, \ell \neq j$:

$$(N_k = N_i) \wedge (X_i = X_j) \wedge (N_j = N_\ell).$$

- $\mathsf{bad}_2 = \mathsf{bad}_{2a} \vee \mathsf{bad}_{2b}$, where:
  - bad_{2a}: $|\{i \in [q_m] : X_i = X_j \wedge N_j = N_k \text{ for some } j \neq i, k \neq j\}| \geq L_1$
  - bad_{2b}: $\sum_{X \in \{0,1\}^{n-1}, d_X > 1} d_X^2 \geq L_2^2$.
- $\mathsf{bad}_3 = \mathsf{bad}_{3a} \vee \mathsf{bad}_{3b} \vee \mathsf{bad}_{3c}$, where:
  - bad_{3a}: $\exists$ a null trail $(i,j) \in [q_m]^{*2}$ of length 2, i.e., $T_i \oplus T_j = \mathbf{0}$.
  - bad_{3b}: $\exists$ a null trail of length 3.
  - bad_{3c}: $\exists$ a null trail of length 4.
- $\mathsf{bad}_4 = \mathsf{bad}_{4a} \vee \mathsf{bad}_{4b}$, where
  - bad_{4a} : $\exists (i,j) \in [q_m] \times [v_m]$ such that $(N_i, X_i, T_i) = (N'_j, X'_j, T'_j)$.
  - bad_{4b} : $\exists (i,j,k,\ell) \in [q_m]^{*3} \times [v_m]$ such that (i,j,k) is an N-trail and

$$(X_k = X'_\ell) \wedge (N'_\ell = N_i) \wedge (T_i \oplus T_j \oplus T_k \oplus T'_\ell = \mathbf{0}).$$

- $\mathsf{bad}_5 = \mathsf{bad}_{5a} \vee \mathsf{bad}_{5b} \vee \mathsf{bad}_{5c} \vee \mathsf{bad}_{5d}$, where:
  - bad_{5a}: $\exists$ a n-collision of length 1 trails.
  - bad_{5b}: $\exists$ a n-collision of length 2 N-trails.
  - bad_{5c}: $\exists$ a n-collision of length 2 X-trails.
  - bad_{5d}: $\exists$ a n-collision of length ≥ 3 trails.
- bad_6 : $q_c \geq \frac{2^{2n}}{186 q_m^2}$.

Interpretations of bad events. We make the following interpretations and implications of the bad events, which are used in the analysis multiple times. The detailed description and analysis are deferred to the end of Section 6.3.

Fact 1 *If* $\neg\mathsf{bad}_1$*, then it holds that*

1. *every length-4 trail is* N*-trail,*
2. $\nexists$ *length-5 trail,*

3. $\nexists$ *cycles in* $\mathcal{G}^{=}(\tau)$*,*
4. $\nexists (i,j) \in [q_m]^{*2}$ *s.t.* $(N_i = N_j) \wedge (X_i = X_j)$*.*

Furthermore, each component $\mathcal{C}$ *of* $\mathcal{G}^{=}(\tau)$ *of size* ≥ 3 *can be understood as a tree, which we call* $\mathrm{tree}_{\geq 3}$*, (See Figure 3) with a special vertex* N_0 *called as a* root*. Every vertices with degree 1 in the tree is called by a* leaf*.*

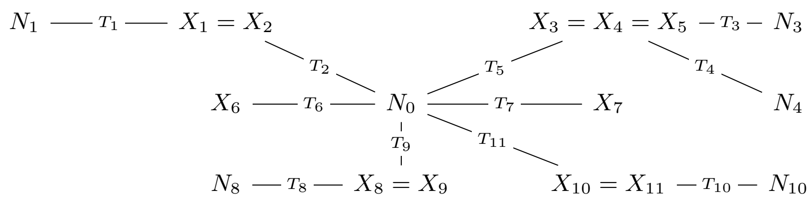

Fig. 3: An example of $\mathrm{tree}_{\geq 3}$. In each edge, the tag T_i corresponds to the query $\mathsf{P}(0\|N_i) \oplus \mathsf{P}(1\|X_i)$, where X_i and N_i are written in each vertex. The root is N_0, which is equal to N_2, N_3, N_6, N_7, N_9, and N_{11}. $N_1, N_3, N_4, X_6, X_7, N_8$, and N_{10} are leaves.

Fact 2 *If* $\neg\mathsf{bad}_1, \neg\mathsf{bad}_3$ *and* $\neg\mathsf{bad}_4$*, then* $\mathcal{G}(\tau)$ *is nice.*

Fact 3 *If* $\neg\mathsf{bad}_1$ *and* $\neg\mathsf{bad}_2$*, the following upper bounds hold:*

- *The number of all vertices in all* $\mathrm{tree}_{\geq 3}$ *is less than or equal to* $3L_1 + \mu_m$*.*
- $d_X \leq L_2$ *for all* $X \in \{0,1\}^{n-1}$ *and* $\xi_{\max} \leq 2L_1 + 2L_2 + \mu_m$*. Furthermore,* $\xi_{\max} q_m \leq \frac{5 \cdot 2^n}{32} \leq \frac{2^n}{4}$ *holds.*
- *The number of length-2* N*-trails is bounded by* $L_2^2/2$*.*
- *The number of length-2* X*-trails is bounded by* $2\mu_m^2$ *(regardless of* bad_2*).*
- *Recall the notations from eq. (4). The number of trails in* $\mathcal{C}_1, \ldots, \mathcal{C}_\alpha$ *is bounded by* $2\mu_m^2 + 9L_1^2 + 0.5L_2^2$*. Further, it holds that* $\sum_{i=1}^{\alpha} c_i^2 \leq 18L_1^2 + L_2^2 + 4\mu_m^2 \leq \min\left(4.5\xi_{\max}^2, \frac{2^n}{16n}\right)$*.*

Fact 4 *If* $\neg\mathsf{bad}_1$ *and* $\neg\mathsf{bad}_3$*, a collision pair of two trails does not start from the same vertex. More strongly, for a* ℓ*-collision* $\{\gamma_1, \ldots, \gamma_\ell\}$*, there exist a set of indices* $\{i_1, \ldots, i_\ell\}$ *such that for each* j*,* i_j *is included in* γ_j *but not included in* γ_k *for all* $k < j$*.*

Fact 5 *If* H *is a* δ*-AXU hash function,* $\mathbf{Ex}[q_c] \leq 2\mu_m + q_m^2\delta$ *and* $\mathbf{Ex}\left[q_c^2\right] \leq 8\mu_m^2 + 2q_m^3\delta$*.*

Bad transcript analysis. The probability $\Pr[\mathsf{bad}]$ is bounded as follows:

$$\epsilon_2 := \frac{\ell(7\mu_m^2 + 2v_m)}{2^n} + \frac{3\ell q_m^2 L_1}{2^{2n}} + \frac{3\ell\mu_m q_m}{2^n L_1} + \frac{3\ell q_m^2}{2^n L_2^2} + \frac{372\ell q_m^4}{2^{3n}}. \tag{32}$$

The detailed analysis is deferred to Section 6.3.

Good transcript analysis. We now assume that the transcript is good, i.e., no bad events occur. Recall $\mathcal{R}_i$ is defined as follows for $i \in [\alpha+\beta]$:

$$\mathcal{R}_i = \left\{(V_1, V_1', V_2, V_2') \in {\mathcal{C}_i}^{*2} \times {\mathcal{C}_j}^{*2} \,\middle|\, j < i \text{ and } \lambda(V_1, V_1') = \lambda(V_2, V_2')\right\},$$

which is a missing term in the above analysis. We divide it into two sets:

$$\mathcal{S}_i := \left\{(V_1, V_1', V_2, V_2') \in \mathcal{R}_i \,\middle|\, \overline{V_1 V_1'}, \overline{V_2 V_2'} \in \mathcal{E}\right\},$$
$$\mathcal{D}_i = \mathcal{R}_i \setminus \mathcal{S}_i.$$

Let $S := \sum_{i=1}^{\alpha+\beta} |\mathcal{S}_i|$. Since $\cup_{i\in[\alpha+\beta]}\mathcal{S}_i$ is the number of collisions of the independent uniform random tags over $\{0,1\}^n \backslash \{\mathbf{0}\}$ among edges, we can invoke Lemma 5 to obtain

$$\mathbf{Ex}[S] \le \frac{q_m^2}{2B}, \quad \mathbf{Ex}\left[S^2\right] \le \begin{cases} \frac{q_m^2}{B} & \text{if } \frac{q_m^2}{2} < B, \\ \frac{q_m^4}{2B^2} & \text{otherwise,} \end{cases} \tag{33}$$

where $B = 2^n - 1$. Also, by $\neg\mathsf{bad}_5$, it holds that $S \le nq_m$.

Let $C := \sum_{i=1}^{\alpha} c_i^2$. Consider $\mathcal{D}_i$ for $i \le \alpha$. For each $(V_1, V_1') \in \mathcal{C}_i^{*2}$, $\neg\mathsf{bad}_5$ asserts that there are at most $4n$ different $(V_2, V_2') \in \mathcal{V}^{*2}$ with the same label with $\lambda(V_1, V_1')$. On the other hand, for $i > \alpha$, a pair of vertices (V_2, V_2') such that $(V_1, V_1', V_2, V_2') \in \mathcal{D}_i$ must be included in $\mathcal{C}_j$ for $j \le \alpha$, because it is included in $\mathcal{S}_i$ otherwise. For each $(V_2, V_2') \in \mathcal{C}_j^{*2}$ for $j \le \alpha$, there are at most n different $i > \alpha$ such that $(V_1, V_1', V_2, V_2') \in \mathcal{D}_i$ for some V_1, V_2. Therefore, we have

$$\sum_{i=1}^{\alpha+\beta} |\mathcal{D}_i| \le 5n \sum_{i=1}^{\alpha} \binom{c_i}{2} \le 3nC.$$

We consider the following upper bound before invoking Theorem 3. From now on, we occasionally give *colors* on some terms to denote the corresponding upper (or lower) bounds in the following (in)equalities, for making one easily chase the transitions of the terms.

$$\begin{aligned}
&\frac{2S + 2(\sum_{i=1}^{\alpha+\beta} |\mathcal{D}_i|) + 2C + 18v_m}{2^n} + \frac{2Cq_c^2 + 31q_c q_m^2}{2^{2n}} + \frac{20q_m^4}{2^{3n}} \\
&\le \frac{2nq_m + 7nC + 18v_m}{2^n} + \frac{9\xi_{\max}^2 \cdot q_m^2 + 31q_c q_m^2}{2^{2n}} + \frac{20q_m^4}{2^{3n}} \\
&\le \frac{1}{128} + \frac{7}{16} + \frac{18}{128} + \frac{225}{32^2} + \frac{20}{8^4} + \frac{31q_m^2\left(\frac{2^{2n}}{186q_m^2}\right)}{2^{2n}} \le 1
\end{aligned}$$

where we use the inequalities from Fact 3 and the upper bounds of q_c from $\neg\mathsf{bad}_6$, $q_m \le \frac{2^{3n/4}}{8} \le \frac{2^n}{12}$. By Theorem 3, it holds that

$$\left|\frac{h(\mathcal{G})(2^n-1)^q}{(2^n)_{|\mathcal{V}|}} - 1\right| \le \frac{4S + 14nC + 36v_m}{2^n} + \frac{4Cq_c^2 + 62q_c q_m^2}{2^{2n}} + \frac{40q_m^4}{2^{3n}} =: \epsilon_1(\tau). \tag{34}$$

6.2 Proof of Theorem 11

We will use Theorem 2 to prove the main theorem in this section given Equations (32) and (34). The remaining part is to give an upper bound of $\epsilon_1(\tau)^2$ to prove the main theorem and optimize the parameters L_1, L_2 appropriately.

By $\neg\mathsf{bad}_6$, Fact 5, Equation (33), and the assumptions on the parameters, especially all terms are less than 1, the expectations of the squared terms can be bound as follows:

$$\begin{aligned}
&\mathbf{Ex}\left[\left(\frac{S}{2^n}\right)^2\right] \leq \frac{q_m^2}{B \cdot 2^{2n}} + \frac{q_m^4}{2B^2 \cdot 2^{2n}} \leq \frac{q_m^4}{2^{3n}} \\
&\mathbf{Ex}\left[\left(\frac{nC}{2^n}\right)^2\right] \leq \mathbf{Ex}\left[\left(\frac{n\xi_{\max} q_c}{2^n}\right)^2\right] \leq \frac{(n(2L_1 + 2L_2 + \mu_m))^2(8\mu_m^2 + 2q_m^3\delta)}{2^{2n}} \\
&\mathbf{Ex}\left[\left(\frac{4Cq_c^2}{2^{2n}}\right)^2\right] \leq \mathbf{Ex}\left[\frac{4Cq_c^2}{2^{2n}}\right] \leq \frac{4(5L_1 + L_2 + 2\mu_m)^2(8\mu_m^2 + 2q_m^3\delta)}{2^{2n}} \\
&\mathbf{Ex}\left[\left(\frac{62q_c q_m^2}{2^{2n}}\right)^2\right] \leq \mathbf{Ex}\left[\frac{62^2 \cdot 8\mu_m^2 q_m^4}{2^{4n}} + \frac{62^2 \cdot 2q_x^2 q_m^4}{2^{4n}}\right] \\
&\leq \mathbf{Ex}\left[\frac{11q_m^4}{2^{3n}} + \frac{42q_x q_m^2}{2^{2n}}\right] \leq \frac{53q_m^4}{2^{3n}}
\end{aligned}$$

In the last inequality, we invoke the notation used in the proof of Fact 5, where $q_c \leq 2\mu_m + q_x$ and $\mathbf{Ex}[q_x] \leq q_m^2\delta$. We also use $\mu_m \leq \frac{2^{0.5n}}{12\sqrt{n}}$ and $q_x \leq q_c \leq \frac{2^{2n}}{186q_m^2}$ by $\neg\mathsf{bad}_6$. We derive an upper bound of $\sqrt{2\mathbf{Ex}[\epsilon_1(\tau)^2]}$ using Lemma 6 as follows:

$$\begin{aligned}
&\frac{29q_m^2}{2^{1.5n}} + \frac{70n(2L_1 + 2L_2 + \mu_m)(4\mu_m^2 + q_m^3\delta)^{0.5}}{2^n} + \frac{125v_m}{2^n} + \frac{139q_m^4}{2^{3n}} \\
&\leq \frac{125v_m + 140n\mu_m^2}{2^n} + \frac{32q_m^2 + 70\ell^{0.5}n\mu_m q_m^{1.5}}{2^{1.5n}} + \frac{140n(L_1 + L_2)(4\mu_m^2 + q_m^3\delta)^{0.5}}{2^n}
\end{aligned}$$

where we use $q_m \leq \frac{2^{3n/4}}{8}$, $2^{n/8} > n$.

Combining with Equation (32), the overall security bound $\sqrt{2u\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]}+2u\epsilon_2$ from Theorem 2 is given by

$$\begin{aligned}
&\frac{14\ell u\mu_m^2+4\ell uv_m}{2^n}+\frac{6\ell uq_m^2L_1}{2^{2n}}+\frac{6\ell u\mu_m q_m}{2^nL_1}+\frac{6\ell uq_m^2}{2^nL_2^2}+\frac{744\ell uq_m^4}{2^{3n}}\\
&+\frac{125\sqrt{u}v_m+140n\sqrt{u}\mu_m^2}{2^n}+\frac{32\sqrt{u}q_m^2+70\ell^{0.5}n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}}\\
&+\frac{314n(L_1+L_2)\sqrt{u}\max\left(\mu_m,q_m^{1.5}\delta^{0.5}\right)}{2^n}\\
\le&\frac{(14\ell u+140n\sqrt{u})\mu_m^2+129\ell uv_m}{2^n}+\frac{60\ell^{0.5}\sqrt{u}q_m^2}{2^{1.5n}}+\frac{70\ell^{0.5}n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}}\\
&+\frac{6\ell uq_m^2L_1}{2^{2n}}+\frac{314nL_1\sqrt{u}\max\left(\mu_m,q_m^{1.5}\delta^{0.5}\right)}{2^n}+\frac{6\ell u\mu_m q_m}{2^nL_1}\\
&+\frac{6\ell uq_m^2}{2^nL_2^2}+\frac{314nL_2\sqrt{u}\max\left(\mu_m,q_m^{1.5}\delta^{0.5}\right)}{2^n}
\end{aligned}$$

where we use $\frac{744\ell uq_m^4}{2^{3n}}\le 1$.

We balance the last equation by choosing

$$L_1^2=\frac{3\ell u\mu_m q_m}{\max(157nu^{0.5}\mu_m,157nu^{0.5}q_m^{1.5}\delta,3q_m^2\delta)},\tag{35}$$

and

$$L_2=\begin{cases}\left(\frac{3\ell u^{0.5}q_m^2}{157n\max(\mu_m,q_m^{1.5}\delta)}\right)^{\frac{1}{3}} & \text{if } q_m^3<2^{2n},\\ \frac{2^n}{32q_m} & \text{if } q_m^3\ge 2^{2n}.\end{cases}\tag{36}$$

We consider two cases separately. If $q_m^3<2^{2n}$, this gives the final advantage upper bound by

$$\begin{aligned}
&\frac{(14\ell u+140n\sqrt{u})\mu_m^2+129\ell uv_m}{2^n}+\frac{60\ell^{0.5}\sqrt{u}q_m^2}{2^{1.5n}}+\frac{70\ell^{0.5}n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}}\\
&+\frac{87\ell^{\frac{1}{2}}n^{\frac{1}{2}}u^{\frac{3}{4}}\mu_m q_m^{0.5}}{2^n}+\frac{87\ell^{\frac{3}{4}}n^{\frac{1}{2}}u^{\frac{3}{4}}\mu_m^{0.5}q_m^{\frac{5}{4}}}{2^{\frac{5n}{4}}}+\frac{12u\mu_m^{0.5}q_m^{1.5}}{2^{1.5n}}\\
&+\frac{87\ell^{\frac{1}{3}}n^{\frac{2}{3}}u^{\frac{2}{3}}\mu_m^{\frac{2}{3}}q_m^{\frac{2}{3}}}{2^n}+\frac{87\ell^{\frac{2}{3}}n^{\frac{2}{3}}u^{\frac{2}{3}}q_m^{\frac{5}{3}}}{2^{\frac{4n}{3}}}\\
\le&\frac{(72\ell u+140n\sqrt{u})\mu_m^2+129\ell uv_m}{2^n}+\frac{61\ell^{0.5}\sqrt{u}q_m^2}{2^{1.5n}}+\frac{70\ell^{0.5}n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}}\\
&+\frac{12u\mu_m^{0.5}q_m^{1.5}}{2^{1.5n}}+\frac{153\ell^{\frac{1}{3}}n^{\frac{2}{3}}u^{\frac{2}{3}}\mu_m^{\frac{2}{3}}q_m^{\frac{2}{3}}}{2^n}+\frac{158\ell^{\frac{2}{3}}n^{\frac{2}{3}}u^{\frac{2}{3}}q_m^{\frac{5}{3}}}{2^{\frac{4n}{3}}}
\end{aligned}$$

where we use $\frac{222\ell u q_m^{2.5}}{2^{2n}} \leq \frac{11.1\ell n u q_m^{2.5}}{2^{2n}} \leq \left(\frac{11.1\ell n u q_m^{2.5}}{2^{2n}}\right)^{2/3}$ and the AM-GM inequality to suppress some terms as follows:

$$\frac{\ell u \mu_m^2}{2^n} + \frac{3\ell^{\frac{1}{3}} n^{\frac{2}{3}} u^{\frac{2}{3}} \mu_m^{\frac{2}{3}} q_m^{\frac{2}{3}}}{2^n} \geq \frac{4\ell^{\frac{1}{2}} n^{\frac{1}{2}} u^{\frac{3}{4}} \mu_m q_m^{\frac{1}{2}}}{2^n},$$
$$\frac{\ell u \mu_m^2}{2^n} + \frac{3\ell^{\frac{2}{3}} n^{\frac{2}{3}} u^{\frac{2}{3}} q_m^{\frac{5}{3}}}{2^{\frac{4n}{3}}} \geq \frac{4\ell^{\frac{3}{4}} n^{\frac{1}{2}} u^{\frac{3}{4}} \mu_m^{\frac{1}{2}} q_m^{\frac{5}{4}}}{2^{\frac{5n}{4}}}.$$

Now we consider the case $q_m^3 \geq 2^{2n}$. First, observe that $1 \leq \frac{q_m^{1.5}}{2^n}$. The overall advantage upper bound becomes:

$$\begin{aligned}
&\frac{(14\ell u + 140n\sqrt{u})\mu_m^2 + 129\ell u v_m}{2^n} + \frac{60\ell^{0.5}\sqrt{u} q_m^2}{2^{1.5n}} + \frac{70\ell^{0.5} n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}} \\
&+ \frac{87\ell^{\frac{1}{2}} n^{\frac{1}{2}} u^{\frac{3}{4}} \mu_m q_m^{0.5}}{2^n} + \frac{87\ell^{\frac{3}{4}} n^{\frac{1}{2}} u^{\frac{3}{4}} \mu_m^{0.5} q_m^{\frac{5}{4}}}{2^{\frac{5n}{4}}} + \frac{12u\mu_m^{0.5} q_m^{1.5}}{2^{1.5n}} \\
&+ \frac{6144\ell u q_m^4}{2^{3n}} + \frac{10n\sqrt{u}\mu_m}{q_m} + \frac{10\ell^{0.5} n\sqrt{u} q_m^{0.5}}{2^{0.5n}} \\
&\leq \frac{(72\ell u + 140n\sqrt{u})\mu_m^2 + 129\ell u v_m}{2^n} + \frac{149\ell^{0.5}\sqrt{u} q_m^2}{2^{1.5n}} + \frac{80\ell^{0.5} n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}} \\
&+ \frac{66\ell^{\frac{1}{3}} n^{\frac{2}{3}} u^{\frac{2}{3}} \mu_m^{\frac{2}{3}} q_m^{\frac{2}{3}}}{2^n} + \frac{66\ell^{\frac{2}{3}} n^{\frac{2}{3}} u^{\frac{2}{3}} q_m^{\frac{5}{3}}}{2^{\frac{4n}{3}}} + \frac{12u\mu_m^{0.5} q_m^{1.5}}{2^{1.5n}}
\end{aligned}$$

where we use the above application of the AM-GM inequality and

$$\frac{n\sqrt{u}\mu_m}{q_m} \leq \frac{n\sqrt{u}\mu_m q_m^{0.5}}{2^n} \leq \frac{n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}}$$
$$\frac{\ell^{0.5} n\sqrt{u} q_m^{0.5}}{2^{0.5n}} \leq \frac{\ell^{0.5} n\sqrt{u} q_m^2}{2^{1.5n}}$$

Taking the maximum of both, we have the advantage upper bound as follows:

$$\begin{aligned}
&\frac{(72\ell u + 140n\sqrt{u})\mu_m^2 + 129\ell u v_m}{2^n} + \frac{149\ell^{0.5}\sqrt{u} q_m^2}{2^{1.5n}} + \frac{80\ell^{0.5} n\sqrt{u}\mu_m q_m^{1.5}}{2^{1.5n}} \\
&+ \frac{153\ell^{\frac{1}{3}} n^{\frac{2}{3}} u^{\frac{2}{3}} \mu_m^{\frac{2}{3}} q_m^{\frac{2}{3}}}{2^n} + \frac{153\ell^{\frac{2}{3}} n^{\frac{2}{3}} u^{\frac{2}{3}} q_m^{\frac{5}{3}}}{2^{\frac{4n}{3}}} + \frac{12u\mu_m^{0.5} q_m^{1.5}}{2^{1.5n}}
\end{aligned}$$

This concludes the concrete security of the main theorem.

Sanity check. Our choices of L_1, L_2 for the optimizations should obey the conditions $L_1, L_2 \ll \min\left(\sqrt{\frac{2^n}{n}}, \frac{2^n}{q_m}\right)$. Recall we choose them according to Equations (35) and (36). Since $\frac{1}{\max(x,y,z)} \leq \min\left(\frac{1}{x}, \frac{1}{y}, \frac{1}{z}\right)$, it suffices to check one of the choice make L_1, L_2 satisfy the condition. For L_1, choosing $nu^{0.5}\mu_m$ among three choices for the maximum gives

$$L_1^2 = O\left(\frac{\ell u^{0.5} q_m}{n}\right)$$

which is much smaller than $\frac{2^n}{n}$ because of the assumption $uq_m^2 \le (2^n/\ell)^2$. Also, choosing $m = nu^{0.5}q_m^{1.5}\delta$ gives

$$L_1^2 = O\left(\frac{\ell^{0.5}u^{0.5}2^{0.5n}}{nq_m^{0.5}}\right)$$

which is smaller than $\left(\frac{2^n}{q_m}\right)^2$. This is because it is equivalent to $\ell u\mu_m^2 q_m^3 \ll n^2 2^{3n}$, which is true because of the condition $\mu_m q_m \ll 2^n$.

For L_2, if $q_m^3 < 2^{2n}$, choosing $q_m^{1.5}\delta^{0.5}$ for the maximum gives

$$L_2^3 = O\left(\frac{\ell u^{0.5}q_m^2}{nq_m^{1.5}\delta^{0.5}}\right) = O\left(\frac{\ell^{0.5}u^{0.5}q_m^{0.5}2^{0.5n}}{n}\right)$$

which is smaller than $\frac{2^{1.5n}}{n^{1.5}}$ because it is equivalent to $\ell nuq_m \ll 2^{2n}$. This is also smaller than $\left(\frac{2^n}{q_m}\right)^3$, which is equivalent to $\ell uq_m^7 \ll n^2 2^{5n}$. This is true because of $uq_m^4 \ll 2^{3n}$. If $q_m^3 \ge 2^{2n}$, $\frac{2^n}{q_m} \ll \frac{2^{0.5n}}{\sqrt{n}}$ apparently holds.

6.3 Bad Transcript Analysis and Interpretations

We give an upper bound of the probability that the event

$$\mathsf{bad} = \mathsf{bad}_1 \vee \mathsf{bad}_2 \vee \mathsf{bad}_3 \vee \mathsf{bad}_4 \vee \mathsf{bad}_5$$

occurs in the ideal world. Recall that μ_m is the upper bound of the number of faulty queries, and H is a δ-AXU hash function for $\delta = \ell/2^n$ and $B = 2^n - 1$.

The following fact can be easily shown by an inductive argument: for $k \ge 1$ and uniform and independent random variables $T_1, ..., T_k$ sampled from $\{0,1\}^n \setminus \{\mathbf{0}\}$, it holds that for any $K \in \{0,1\}^n$

$$\Pr\left[\bigoplus_{i\in[k]} T_i = K\right] \le \frac{1}{2^n - 1} = \frac{1}{B}. \tag{37}$$

We now analyze the probability that each bad event occurs. We assume that $n \ge 20$, and $\mu_m, L_1, L_2 \ge 1$. The detailed conditions on the parameters will be explicitly described after analyzing each bad event.

bad_1: The number of indices $i \in [q_m]$ such that there exists $k(\ne i) \in [q_m]$ such that $N_i = N_k$ is bounded by $2\mu_m$. Thus, there are at most $4\mu_m^2$ pairs of $(i,j) \in [q_m]^{*2}$ satisfying the condition. For each (i,j), the probability that $X_i = X_j$, or equivalently $\mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) = N_i \oplus N_j$ is at most δ because H is a δ-AXU. By the union bound, we have

$$\Pr[\mathsf{bad}_1] \le 4\mu_m^2\delta.$$

bad_{2a}: Fix $i \in [q_m]$. There are at most $2\mu_m$ choices of j since it is a repeated nonce. For each j, the probability that $X_i = X_j$ is at most δ, and the probability that i satisfies the condition is at most $2\mu_m\delta$. Therefore, the expected size of the given set is at most $2\mu_m q_m \delta$, and by Markov's inequality, we have

$$\Pr[\mathsf{bad}_{2a}] \leq \frac{2\mu_m q_m \delta}{L_1}.$$

bad_{2b}: Recall the graph-theoretic interpretation; d_X is the number of indices $i \in [q_m]$ such that $X_i = X$. Let Col be the number of $i < j$ such that $X_i = X_j$, whose expectation is less than $q_m^2\delta/2$ because of the δ-AXU property of H. On the other hand, it holds that

$$\mathsf{Col} = \sum_{X \in \{0,1\}^{n-1}} \binom{d_X}{2} \geq \sum_{X : d_X > 1} d_X^2/4,$$

where we use $d_X - 1 \geq d_X/2$ for $d_X > 1$. By Markov's inequality, we have

$$\Pr[\mathsf{bad}_{2b}] = \Pr\left[\sum_{X : d_X > 1} d_X^2 \geq L_2^2\right] \leq \Pr[4\mathsf{Col} \geq L_2^2] \leq \frac{2q_m^2\delta}{L_2^2}.$$

bad_{3a}: Assume that $\neg\mathsf{bad}_1$ so that $i \neq j$ satisfies at most one of $N_i = N_j$ or $X_i = X_j$. We consider the following two cases: 1) $T_i = T_j$ and $N_i = N_j$: The number of pairs (i,j) such that $N_i = N_j$ is at most $2\mu_m^2$ (Fact 3), and the probability that $T_i = T_j$ is $1/B$. 2) $T_i = T_j$ and $X_i = X_j$: For each i,j, the probability that $X_i = X_j$ is bounded by δ, and $T_i = T_j$ is $1/B$ and two events are independent. By the union bound, we have

$$\Pr[\mathsf{bad}_{3a} | \neg\mathsf{bad}_1] \leq \frac{2\mu_m^2 + \delta q_m^2}{B}.$$

bad_{3b}: Suppose that there exist indices (i,j,k) such that $N_i = N_j$ and $X_j = X_k$. The number of (i,j) such that $N_i = N_j$ is at most $2\mu_m^2$. For each k, the two events that $X_j = X_k$ and $T_k = T_i \oplus T_j$ are independent, and the probabilities for them are bounded by δ and $1/B$. By the union bound, we have

$$\Pr[\mathsf{bad}_{3b}] \leq \frac{2\delta\mu_m^2 q_m}{B}.$$

bad_{3c}: Assume that $\neg\mathsf{bad}_1$ and $\neg\mathsf{bad}_{2a}$. By Fact 1, the length-4 trail (i,j,k,ℓ) must be N-trail, i.e., $X_i = X_j$, $N_j = N_k$, and $X_k = X_\ell$ holds. For each pair $(i,j) \in [q_m]^{*2}$, the probability that $X_i = X_j$ is bounded above by δ due to H. Observe that (ℓ,k,j) satisfies $X_\ell = X_k$ and $N_k = N_j$, which makes at most L_1 different choices of ℓ. Because of the structure of the graph (Figure 3), k is deterministic for given (i,j,ℓ). Since the probability that (i,j,k,ℓ) becomes a null trail is at most $1/B$, by the union bound, we have

$$\Pr[\mathsf{bad}_{3c}] \leq \frac{q_m^2\delta L_1}{B}.$$

bad_{4a}: Assume that $\neg\mathsf{bad}_{3a}$. It implies that there is no $i \neq k$ such that $N_i = N_k$ and $T_i = T_k$. For each verification query (N'_j, M'_j, T'_j), there is at most one MAC query (N_i, M_i, T_i) such that $N_i = N'_j$ and $T_i = T'_j$ holds. For such a pair (i, j), the probability that $X_i = X'_j$ is bounded above by δ because of H. By the union bound, we have

$$\Pr[\mathsf{bad}_{4a} | \neg\mathsf{bad}_{3a}] \leq v_m \delta.$$

bad_{4b}: We assume $\neg\mathsf{bad}_1$ and $\neg\mathsf{bad}_2$. Let $(N'_\ell, M'_\ell, T'_\ell)$ be a verification query included in a length-4 cycle described in the condition. For any (i, j, k, ℓ) satisfying the condition, (i, j, k) should be a length-3 N-trail and N_i should be a leaf with the root N_j because of Fact 1. Thus, given a fixed ℓ, there is a unique pair $(i, j) \in [q_m]^{*2}$ such that $N_i = N'_\ell$ and $X_i = X_j$ holds and (i, j, k^*) becomes a trail for some k^* (otherwise violating Fact 1). Fix (ℓ, i, j). For each $k \in [q_m]$, the probability that $X_k = X'_\ell$ and $T_i \oplus T_j \oplus T_k \oplus T'_\ell = 0^n$ are independent and at most δ and $1/B$, respectively. Therefore, regardless of $N_j = N_k$, we have the following bound using the union bound:

$$\Pr[\mathsf{bad}_{4b} | (\neg\mathsf{bad}_1) \wedge (\neg\mathsf{bad}_2)] \leq \frac{v_m q_m \delta}{B}.$$

bad_{5a}: Since the values T_i are independent of each other in the ideal world and there are $\binom{q_m}{n}$ different pairs $(i_1, ..., i_n)$, we have

$$\Pr[\mathsf{bad}_{5a}] \leq \frac{\binom{q_m}{n}}{B^{n-1}} \leq \left(\frac{q_m}{B}\right)^n$$

where we used $n! \geq 2^n - 1$ for $n \geq 4$ in the middle.

bad_{5b}: Assume that $\neg\mathsf{bad}_1, \neg\mathsf{bad}_2$ and $\neg\mathsf{bad}_3$. Define $\mathcal{B}$ be a set of all collections of different n trails of length 2. By Fact 3, we have

$$|\mathcal{B}| \leq \binom{L_2^2/2}{n} \leq \frac{L_2^{2n}}{B}$$

and we can show that $\Pr[T_{i_j} \oplus T_{i'_j} = T_{i_1} \oplus T_{i'_1}] \leq 1/B$ for each (j, j') by Equation (37) and Fact 4. It gives

$$\Pr[\mathsf{bad}_{5b} | \neg\mathsf{bad}_{1,2,3}] \leq \frac{|\mathcal{B}|}{B^{n-1}} \leq \left(\frac{L_2^2}{B}\right)^n.$$

bad_{5c}: Assume that $\neg\mathsf{bad}_1$,$\neg\mathsf{bad}_2$ and $\neg\mathsf{bad}_3$. A similar argument shows that

$$\Pr[\mathsf{bad}_{5c} | \neg\mathsf{bad}_{1,2,3}] \leq \left(\frac{2\mu_m^2}{B}\right)^n.$$

bad_{5d}: Assume that $\neg\mathsf{bad}_1$,$\neg\mathsf{bad}_2$, and $\neg\mathsf{bad}_3$. Each trail of length ≥ 3 should be included in $\mathcal{C}_1, ..., \mathcal{C}_\alpha$. Using Fact 3, A similar argument shows that

$$\Pr[\mathsf{bad}_{5d} | \neg\mathsf{bad}_{1,2,3}] \leq \left(\frac{2\mu_m^2 + 9L_1^2 + 0.5L_2^2}{B}\right)^n.$$

bad_6**:** Recall Fact 5. For $t > 2\mu_m$, by Markov's inequality, it holds that

$$\Pr[q_c \geq 2t] \leq \Pr[q_c - 2\mu_m \geq t] \leq \frac{q_m^2 \delta}{t}.$$

By setting $t = \frac{2^{2n}}{372 q_m^2}$, we have

$$\Pr[\mathsf{bad}_6] \leq \frac{372 q_m^4 \delta}{2^{2n}}.$$

Summary. Recall that $q_m \leq \min\left(\frac{2^n}{12n}, \frac{2^{3n/4}}{4}\right)$, $v_m \leq \frac{2^n}{127}$, and $\mu_m \leq \frac{\sqrt{2^n}}{12\sqrt{n}}$ holds. We will choose $L_1, L_2 \leq \min\left(\frac{2^n}{32 q_m}, \frac{2^{0.5n}}{24\sqrt{n}}\right)$. This setting makes $\Pr[\mathsf{bad}_5] \leq 1/2^n$ and the condition of bad_6 holds. The overall upper bound of $\Pr[\mathsf{bad}]$ is as follows:

$$4\mu_m^2\delta + \frac{2\mu_m q_m \delta}{L_1} + \frac{2q_m^2\delta}{L_2^2} + \frac{2\mu_m^2 + (q_m + 2\mu_m^2 + v_m) q_m \delta}{B}$$
$$+ v_m\delta + \frac{q_m^2 \delta L_1}{B} + \frac{1}{2^n} + \frac{372 q_m^4 \delta}{2^{2n}}.$$

Using $\delta = \ell/2^n$ for $\ell \geq 1$, $B = 2^n - 1 \geq 1.0001 \cdot 2^n$, and $q_m \leq 0.01 \cdot 2^n, 2^{3n/4}/8$, we derive the following simplified upper bound:

$$\frac{\ell(7\mu_m^2 + 2v_m)}{2^n} + \frac{3\ell q_m^2 L_1}{2^{2n}} + \frac{3\ell \mu_m q_m}{2^n L_1} + \frac{3\ell q_m^2}{2^n L_2^2} + \frac{372 \ell q_m^4}{2^{3n}}.$$

Analysis of Interpretations. We give detailed descriptions for the interpretations of the bad events. Remind that $\mathcal{G}$ is a bipartite graph.

Fact 1 Suppose that (i, j, k, ℓ) is a length-4 X-trail. Then

$$N_i = N_j, X_j = X_k, N_k = N_\ell$$

holds, which directly violates $\neg\mathsf{bad}_1$. Since a length-5 trail must contain a length-4 X-trail, the second item follows. By this observation, a cycle in $\mathcal{G}^{=}(\tau)$ must be of length 4, which again violates $\neg\mathsf{bad}_1$ if it exists. The final item is just $\neg\mathsf{bad}_1$ with $k = j, \ell = i$. The structure of the graph directly follows.

Fact 2 $\neg\mathsf{bad}_1$ and $\neg\mathsf{bad}_3$ implies that $\mathcal{G}^{=}$ is acyclic and non-degenerated, respectively. The consistency between $\mathcal{G}^{=}$ and $\mathcal{G}^{\neq}$ is due to $\neg\mathsf{bad}_4$, because $\neg\mathsf{bad}_1$ already rules out the other cases.

Fact 3 $\neg\mathsf{bad}_1$ imposes the structure as in Figure 3. The number of vertices included in the trail of length two from the root to the leaf is bounded by $3L_1$ because of $\neg\mathsf{bad}_2$; in particular, if every length of two trails is in the same component, we can count it as $2L_1+1$. The indices corresponding to the other vertices must induce nonce collisions, thus the number of the other vertices is bounded above by μ_m.

For the second item, $d_X \leq L_2$ is obvious from $\neg\mathsf{bad}_2$. For giving an upper bound of $\xi_{\max}$, let us consider the component with a trail of length three. Then, as in the above argument, this component has at most $2L_1+\mu_m+1$ vertices. On the other hand, for the component without such a trail, it must be a star-shape with indices $i_1, ..., i_k$ such that $N_{i_1} = ... = N_{i_k}$ or $X_{i_1} = ... = X_{i_k}$. In any case, the number of vertices in this component is bounded by $\max(\mu_m, L_2) + 1$, all of which are less than $2L_1 + 2L_2 + \mu_m$. For the number of length-2 N-trails, it is at most $\sum_{d_X \geq 2} \binom{d_X}{2} \leq L_2^2/2$. The number of $i \in [q_m]$ with $N_j = N_i$ for some $j \neq i$ is at most $2\mu_m$ and the number of such j is at most μ_m, thus the total number is at most $2\mu_m^2$.

Finally, the number of all trails in $\mathcal{C}_1, ..., \mathcal{C}_\alpha$ is bounded by $\binom{3L_1+\mu_m}{2} + L_2^2/2$ following the above facts. The upper bound of $\sum_{i=1}^{\alpha} c_i^2$ is obvious.

Fact 4 Given a collision pair shares the same starting vertex, we can construct a null trail by combining them and removing the intersection, violating $\neg\mathsf{bad}_3$. We can choose each starting vertex for the second statement as a unique index.

Fact 5 For each $i \in [q_m]$, let I_i equal 1 if there exists $j \in [q_m]$ such that $i \neq j$ and $X_i = X_j$, and zero otherwise. It holds that $\mathbf{Ex}[I_i] \leq q_m\delta$ by the union bound. Let $q_x = \sum_{i\in[q_m]} I_i \leq q_m$. It holds that $q_c \leq 2\mu + q_x$, where 2μ is from the faulty queries, and $q_c^2 \leq (2\mu + q_x)^2 \leq 8\mu_m^2 + 2q_x^2 \leq 8\mu_m^2 + 2q_cq_x$. We have

$$\mathbf{Ex}[q_c] \leq 2\mu + \mathbf{Ex}[q_x] \leq 2\mu + \sum_{i\in[q_m]} \mathbf{Ex}[I_i] \leq 2\mu + q_m^2\delta,$$
$$\mathbf{Ex}\left[q_c^2\right] \leq \mathbf{Ex}\left[8\mu_m^2 + 2q_cq_x\right] = 8\mu_m^2 + 2q_m\mathbf{Ex}[q_x] \leq 8\mu_m^2 + 2q_m^3\delta.$$

6.4 Nonce-respecting Setting

We roughly sketch the security analysis of nEHtM when we only consider the nonce-respecting setting; every constant factor is ignored here. In this case, we can ignore the events $\mathsf{bad}_1, \mathsf{bad}_{2a}$, most of the cases of bad_3 (except the length-2 null trail with the $X_i = X_j$ case), bad_{5c} and bad_{5d}. Asymptotically, the remaining probability of the bad events is

$$\epsilon_2 = O\left(\frac{\ell v_m}{2^n} + \frac{\ell q_m^2}{2^n L_2^2} + \frac{\ell q_m^4}{2^{3n}}\right),$$

where we ignore the probability of bad_5, which can be made less than $1/2^{3n}$.

Since there is no faulty query, the parameter L_2 provides an upper bound of $\xi_{\max} = O(L_2)$ and $C = O(L_2^2)$ as well. We have

$$\begin{aligned}&\frac{S + (\sum_{i=1}^{\alpha+\beta} |\mathcal{D}_i|) + C + v_m}{2^n} + \frac{Cq_c^2 + q_cq_m^2}{2^{2n}} + \frac{q_m^4}{2^{3n}}\\ &= \frac{nq_m + nL_2^2 + v_m}{2^n} + \frac{L_2^2q_c^2 + q_cq_m^2}{2^{2n}} + \frac{q_m^4}{2^{3n}} = O(1).\end{aligned}$$

Let

$$\epsilon_1(\tau) = \frac{S + nC + v_m}{2^n} + \frac{Cq_c^2 + q_cq_m^2}{2^{2n}} + \frac{q_m^4}{2^{3n}}.$$

Then, as in the original proof, we have

$$\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]^{\frac{1}{2}} = O\left(\frac{\ell^{0.5}q_m^2}{2^{1.5n}} + \frac{\ell^{0.5}nL_2q_m^{1.5}}{2^{1.5n}} + \frac{v_m}{2^n}\right).$$

Therefore, by Theorem 2, we have the overall asymptotic advantage bound of $\mathbf{Ex}\left[2u\epsilon_1(\tau)^2\right]^{\frac{1}{2}} + 2u\epsilon_1$ is given by

$$O\left(\frac{\ell^{0.5}\sqrt{u}q_m^2}{2^{1.5n}} + \frac{\ell^{0.5}nL_2\sqrt{u}q_m^{1.5}}{2^{1.5n}} + \frac{\ell uv_m}{2^n} + \frac{\ell uq_m^2}{2^nL_2^2} + \frac{\ell uq_m^4}{2^{3n}}\right).$$

By choosing $L_2 = \Theta\left(\min\left(\left(\frac{\ell uq_m2^n}{n^2}\right)^{\frac{1}{6}}, \frac{2^n}{q_m}\right)\right)$, we have

$$O\left(\frac{\ell uv_m}{2^n} + \left(\frac{\ell uq_m^4}{2^{3n}}\right)^{\frac{1}{2}} + \left(\frac{\ell^2n^2u^2q_m^5}{2^{4n}}\right)^{\frac{1}{3}}\right), \tag{38}$$

where we suppress some terms, as in the main proof. The sanity check passes in exactly the same way.

6.5 Using Stronger Hash and Proofs in [11]

We briefly analyze the security of nEHtM when H satisfies a stronger property as follows: For $\delta > 0$, we say that $\mathsf{H} : \mathcal{K} \times \mathcal{M} \to \mathcal{X}$ is a *pairwise* δ-almost XOR universal, denoted by δ-AXU$^{(2)}$, hash function if it is a δ-AXU and additionally for any $M_1 \neq M_1'$ and $M_2 \neq M_2'$ in $\mathcal{M}$ such that $\{M_1, M_1'\} \neq \{M_2, M_2'\}$ and $X_1, X_2 \in \mathcal{X}$, it holds that

$$\Pr_{K \xleftarrow{\$} \mathcal{K}}\left[\mathsf{H}_{K_h}(M_1) \oplus \mathsf{H}_{K_h}(M_1') = X_1 \wedge \mathsf{H}_{K_h}(M_2) \oplus \mathsf{H}_{K_h}(M_2') = X_2\right] \leq \delta^2.$$

Note that a 4-wise $\ell/|\mathcal{X}|$-almost universal hash function for constant ℓ is a pairwise δ-AXU. We let $\delta = \tilde{O}(1/2^n)$ and ignore the small factors in the analysis for exhibiting the asymptotic behavior solely.

This new property of hash functions allows multiple improvements in our analysis. We first see a variant of Fact 5, which was erroneously used in [11] without any clarification or explicitly using δ-AXU$^{(2)}$.

Fact 6 *If* H *is* δ*-AXU*$^{(2)}$*, then* $\mathbf{Ex}\left[q_c^2\right] = O\left(\mu_m^2 + \frac{q_m^2}{2^n} + \frac{q_m^4}{2^{2n}}\right)$.

Proof (sketch). Recall the definitions in the proof of Fact 5 (at the end of Section 6.3). Let $I_{i,j}$ equal 1 if $X_i = X_j$ and otherwise zero. Then $I_i = \vee_{j\neq i} I_{i,j} \leq \sum_j I_{i,j} = \tilde{O}(q_m/2^n)$. The δ-AXU$^{(2)}$ property implies that $\mathbf{Ex}\left[I_{i,j}I_{k,\ell}\right] \leq \delta^2 = \tilde{O}(1/2^{2n})$. We can give an upper bound of $\mathbf{Ex}\left[q_x^2\right] = \mathbf{Ex}\left[\sum_{i,j} I_iI_j\right]$ by

$$\mathbf{Ex}\left[\sum_i I_i\right] + \mathbf{Ex}\left[\left(\sum_{i,k} I_{i,k}\right)\left(\sum_{j\neq i,\ell} I_{j,\ell}\right)\right] = \tilde{O}\left(\frac{q_m^2}{2^n} + \frac{q_m^4}{2^{2n}}\right).$$

Finally, $q_c^2 \leq (2\mu_m + q_x)^2 \leq 8\mu_m^2 + 2q_x^2$ gives the desired result. □

We take a similar approach whenever two XOR equations of H appear.

- Applying Chebyshev inequality instead of Markov, $\Pr[\mathsf{bad}_{2a}] = \tilde{O}\left(\frac{\mu_m q_m}{2^n L_1^2}\right)$. This requires $L_1 \gg \frac{\mu_m q_m}{2^n}$, and our choice satisfies this constraint.
- We can directly give a better bound $\Pr[\mathsf{bad}_{3c}] = \tilde{O}\left(\frac{\mu_m^2 q_m^2}{2^{3n}}\right)$.
- By Chebyshev, $\Pr[\mathsf{bad}_6] = O\left(\frac{q_m^6}{2^{5m}} + \frac{q_m^8}{2^{6m}}\right)$.

This gives the following upper bound of ϵ_2.

$$\tilde{O}\left(\frac{\mu_m^2 + v_m}{2^n} + \frac{\mu_m q_m}{L_1^2 2^n} + \frac{q_m^2}{L_2^2 2^n} + \frac{q_m^8}{2^{6m}}\right),$$

where we suppress the terms by the AM-GM inequality. For example, $\frac{2q_m^6}{2^{5n}} \leq \frac{q_m^4}{2^{4n}} + \frac{q_m^8}{2^{6n}}$.

We also have better bounds for computing $\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]$ using Fact 6.

$$\begin{aligned}
&\mathbf{Ex}\left[\left(\frac{nC}{2^n}\right)^2\right] = \tilde{O}\left(\frac{Cq_c^2}{2^{2n}}\right)\\
&= \tilde{O}\left((L_1+L_2)^2\left(\frac{\mu_m^2}{2^{2n}} + \frac{q_m^2}{2^{3n}} + \frac{q_m^4}{2^{4n}}\right) + \frac{\mu_m^4}{2^{2n}} + \frac{\mu_m^2 q_m^2}{2^{3n}} + \frac{\mu_m^2 q_m^4}{2^{4n}}\right),\\
&\mathbf{Ex}\left[\left(\frac{Cq_c^2}{2^{2n}}\right)^2\right] = \tilde{O}\left(\frac{Cq_c^2}{2^{2n}}\right),\\
&\mathbf{Ex}\left[\left(\frac{q_c q_m^2}{2^{2n}}\right)^2\right] = \tilde{O}\left(\frac{\mu_m^2 q_m^4}{2^{4n}} + \frac{q_m^6}{2^{5n}} + \frac{q_m^8}{2^{6n}}\right).
\end{aligned}$$

We have an asymptotic upper bound of $\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]^{\frac{1}{2}}$ by

$$\tilde{O}\left(\frac{\mu_m^2}{2^n} + \frac{\mu_m q_m}{2^{1.5n}} + \frac{\mu_m q_m^2}{2^{2n}} + \frac{q_m^4}{2^{3n}} + (L_1+L_2)\left(\frac{\mu_m}{2^n} + \frac{q_m}{2^{1.5n}} + \frac{q_m^2}{2^{2n}}\right)\right)$$

Let $\left(\frac{\mu_m}{2^n} + \frac{q_m}{2^{1.5n}} + \frac{q_m^2}{2^{2n}}\right) =: \nu$. The asymptotic advantage upper bound becomes

$$\begin{aligned}
&\frac{u\mu_m^2 + uv_m}{2^n} + \frac{u\mu_m q_m}{L_1^2 2^n} + \frac{uq_m^2}{L_2^2 2^n} + \frac{uq_m^8}{2^{6m}}\\
&+ \frac{\sqrt{u}\mu_m^2}{2^n} + \frac{\sqrt{u}\mu_m q_m}{2^{1.5n}} + \frac{\sqrt{u}\mu_m q_m^2}{2^{2n}} + \frac{\sqrt{u}q_m^4}{2^{3n}} + \sqrt{u}(L_1+L_2)\nu\\
&\lesssim \frac{u\mu_m^2 + uv_m}{2^n} + \frac{u\mu_m q_m}{L_1^2 2^n} + \frac{uq_m^2}{L_2^2 2^n} + \frac{\sqrt{u}q_m^4}{2^{3n}} + \sqrt{u}(L_1+L_2)\nu
\end{aligned}$$

If $q_m^3 < 2^{2n}$, taking $L_1^3 = \frac{u^{0.5}\mu_m q_m}{2^n \nu}, L_2^3 = \frac{u^{0.5} q_m^2}{2^n \nu}$ gives the asymptotic advantage:

$$\tilde{O}\left(\frac{u\mu_m^2 + uv_m}{2^n} + \frac{\sqrt{u}q_m^4}{2^{3n}} + \frac{u^{\frac{2}{3}}\mu_m^{\frac{2}{3}} q_m^{\frac{2}{3}}}{2^n} + \frac{u^{\frac{2}{3}} q_m^{\frac{4}{3}}}{2^{\frac{4n}{3}}} + \frac{u^{\frac{2}{3}} q_m^2}{2^{\frac{5n}{3}}}\right)$$

Otherwise, if $q_m^3 \le 2^{2n}$, we can choose $L_2 = \Theta\left(\frac{2^n}{q_m}\right)$ instead, giving the same upper bound. If $\mu_m > 0$, $\frac{2u^{\frac{2}{3}} q_m^{\frac{4}{3}}}{2^{\frac{4n}{3}}} \le \frac{u^{\frac{2}{3}} q_m^{\frac{2}{3}}}{2^n} + \frac{u^{\frac{2}{3}} q_m^2}{2^{\frac{5n}{3}}}$ gives a simpler bound. When $\mu_m = v_m = 0$, we have the multi-user security of nEHtM as pseudorandom functions with the punctured codomain $\{0,1\}^n \setminus \{\mathbf{0}\}$:

$$\tilde{O}\left(\frac{\sqrt{u}q_m^4}{2^{3n}} + \left(\frac{uq_m^2}{2^{2n}}\right)^{2/3} + \left(\frac{u^2 q_m^6}{2^{5n}}\right)^{1/3}\right). \tag{39}$$

Recovering the result of [11] using δ-AXU$^{(2)}$. We give the correct asymptotic bound of the multi-use $\mathsf{nEHtM2}$ security following the proof of [11] assuming that H is δ-AXU$^{(2)}$. Recall the following bound from [11, eprint version, page 27], which are based on the slightly worse version of Theorem 3 with a different bad event for bounding q_c.

$$\epsilon_1(\tau) = \tilde{O}\left(\frac{\sum_{i=1}^{\alpha+\beta} |\mathcal{S}_i|}{2^n} + \frac{Lq_c}{2^n} + \frac{Lq_c q_m^2}{2^{2n}} + \frac{Lq_m^4}{2^{3n}}\right),$$

$$\epsilon_2 = \tilde{O}\left(\frac{q_m^2}{2^{2n}} + \frac{q_m^2}{L^2 2^n} + \frac{L^2 q_m^8}{2^{6n}}\right),$$

where $\tilde{O}$ ignores the polynomial of n and ℓ. We also remove some terms in ϵ_2, which only makes the bound better. A straightforward computation using Fact 6 (i.e., assuming H is δ-AXU$^{(2)}$) gives the following bound.

$$\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]^{1/2} = \tilde{O}\left(\frac{Lq_m}{2^{1.5n}} + \frac{Lq_m^4}{2^{3n}}\right).$$

We suppress most of the terms using the AM-GM inequality and $L \ge 1$. We stress that the red term in the above bound is missing in the final bound of [11, Theorem 5], which corresponds to

$$\frac{4\sqrt{2u}(n+1)Lq_m\delta^{1/2}}{2^n}$$

in their notation and appeared in the second line of $\mathsf{Adv}^{\mathsf{mu-prf}}_{\mathsf{nEHtM}}(u, q_{\max})$ of page 28. This makes the actual security bound slightly worse than they claimed, even assuming that H is a pairwise δ-AXU. Taking $L^3 = \sqrt{u}\min\left(2^{0.5n}q_m, \frac{2^{2n}}{q_m^2}\right)$ gives the final bound of

$$\tilde{O}\left(\left(\frac{u^2 q_m^4}{2^{4n}}\right)^{1/3} + \left(\frac{u^2 q_m^{10}}{2^{7n}}\right)^{1/3}\right).$$

The second term indeed appears in the original statement, and the first term is larger than $\frac{uq_m^2}{2^{2n}}$ in the original bound. Also, the following inequality confirms that the dominating term $\left(\frac{u^2 q_m^6}{2^{5n}}\right)^{1/3}$ in the original bound is just hidden by the other terms; i.e., our analysis does not miss the term.

$$3\left(\frac{u^2 q_m^6}{2^{5n}}\right)^{1/3} \leq 2\left(\frac{u^2 q_m^4}{2^{4n}}\right)^{1/3} + \left(\frac{u^2 q_m^{10}}{2^{7n}}\right)^{1/3}$$

Finally, our new bound in Equation (39) with the same assumption is always tighter than this bound, because

$$\begin{cases} \left(\frac{u^2 q_m^4}{2^{4n}}\right)^{1/3} \text{ is the dominating term} & \text{if } q_m^2 \leq 2^n \text{, and} \\ \left(\frac{u^2 q_m^{10}}{2^{7n}}\right)^{1/3} \geq \left(\frac{u^2 q_m^6}{2^{5n}}\right)^{1/3} & \text{if } q_m^2 \geq 2^n. \end{cases}$$

Recovering the result of [11] without δ-AXU$^{(2)}$. If we are willing to avoid δ-AXU$^{(2)}$, we only can use Fact 5, and the probability for bad_6 (denoted by bad_5 in the original paper) becomes worse:

- If we use Markov inequality as in our analysis, $\Pr[\mathsf{bad}_6] = \tilde{O}\left(\frac{Lq_m^4}{2^{3n}}\right)$.
- If we use Chebyshev inequality, $\Pr[\mathsf{bad}_6] = \tilde{O}\left(\frac{L^2 q_m^7}{2^{5n}}.\right)$

Based on this, we can obtain the following asymptotic bounds using Fact 5 only assuming that H is δ-AXU:

$$\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]^{1/2} = \tilde{O}\left(\frac{Lq_m^{1.5}}{2^{1.5n}} + \frac{Lq_m^4}{2^{3n}}\right), \quad \epsilon_2 = \tilde{O}\left(\frac{q_m^2}{2^{2n}} + \frac{q_m^2}{L^2 2^n} + \Pr[\mathsf{bad}_6]\right),$$

where the red terms are worse than the bound assuming δ-AXU$^{(2)}$.

If we use Markov inequality, we take $L^3 = \min\left(\sqrt{u}2^{0.5n}q_m^{0.5}, \frac{\sqrt{u}2^{2n}}{q_m^2}, \frac{2^{2n}}{q_m^2}\right)$ to obtain the final bound of

$$\tilde{O}\left(\left(\frac{u^2 q_m^5}{2^{4n}}\right)^{1/3} + \left(\frac{u^3 q_m^{10}}{2^{7n}}\right)^{1/3}\right).$$

If we use Chebyshev inequality, we take $L^3 = \min\left(\sqrt{u}2^{0.5n}q_m^{0.5}, \frac{\sqrt{u}2^{2n}}{q_m^2}\right)$ or $L^4 = \frac{2^{4n}}{q_m^5}$, and obtain the final bound of

$$\tilde{O}\left(\left(\frac{u^2 q_m^5}{2^{4n}}\right)^{1/3} + \left(\frac{u^2 q_m^{10}}{2^{7n}}\right)^{1/3} + \frac{uq_m^{4.5}}{2^{3n}}\right)$$

In any case, the bound becomes worse than one using δ-AXU$^{(2)}$.

7 Multi-User Security of DbHtS

This section proves the multi-user MAC security of the Double-block Hash-then-Sum (DbHtS) scheme proposed by [20] with the domain separation.

Let $\mathcal{M} = \{0,1\}^*$ be a message space, $\mathcal{K}_h = \{0,1\}^k$ be a hash key space, and $\mathcal{K} = \{0,1\}^k$ be a block cipher key space. Note that we assume $\mathcal{K}_h = \mathcal{K}$ for ease representation. Let $\mathsf{H} = (\mathsf{H}^1, \mathsf{H}^2) : \mathcal{K}_h \times \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^{n-1} \times \{0,1\}^{n-1}$ be a hash function with $(2n-2)$-bit outputs, which can be decomposed into two $(n-1)$-bit hash functions $\mathsf{H}^1, \mathsf{H}^2 : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^{n-1}$ so that $\mathsf{H}_{K_h}(M) = (\mathsf{H}^1_{K_{h_1}}(M), \mathsf{H}^2_{K_{h_2}}(M))$ where $K_h = (K_{h_1}, K_{h_2}) \in \mathcal{K}_h \times \mathcal{K}_h$. Let $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher modeled as an ideal cipher, i.e., keyed random permutations. We define the DbHtS constructions with the domain separation as follows:

$$\mathsf{DbHtS}[\mathsf{H}, \mathsf{E}](K_h, K, M) \stackrel{\text{def}}{=} \mathsf{E}_K(0 \| \mathsf{H}^1_{K_{h,1}}(M)) \oplus \mathsf{E}_K(1 \| \mathsf{H}^2_{K_{h,2}}(M)).$$

It is well-known that MAC security of deterministic MACs can be viewed as PRF security. Using the same reasoning, it is enough to show that PRF* security of DbHtS. We also introduce the additional parameter q_m denoting the maximum number of queries each user makes and assume $q = uq_m$ for our security analysis; this does not lose the generality by making some redundant queries at the end.

Theorem 12 shows the multi-user DbHtS security bound improved from [37] (Recall Figure 1 for the comparison). Following the original paper, we require the hash functions H^1, H^2 used in DbHtS to be regular and AU, and E is implemented by the ideal cipher. The proof can be found in Section 7.1.

Theorem 12 (Proof is in Section 7.1). *Let n, k, u, p, l, and q_m be positive integers such that $p + uq_m l \leq 2^{n-2}$. Let hash functions H^1, $\mathsf{H}^2 : \{0,1\}^k \times \mathcal{M} \to \{0,1\}^{n-1}$ are δ_1-regular and δ_2-AU. Let the block cipher $\mathsf{E} : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ be modeled as an ideal cipher. Let l be the maximum block length among all construction queries. Then, it holds that*

$$\begin{aligned}\mathsf{Adv}^{\mathsf{mu\text{-}prf}^*}_{\mathsf{DbHtS}}(u, q_m, p) \leq\ & \frac{2u}{2^k} + \frac{2uq_m p\delta_1}{2^k} + \frac{4u^2 q_m^2 l\delta_1}{2^k} + \frac{8uq_m^3(\delta_1+\delta_2)}{2^n} + \frac{2uq_m pl}{2^{k+n}} \\ &+ \frac{8uq_m p}{2^{k+n}} + \frac{u^2}{2^{k+n}} + \frac{u(3u+p)(6u+2p)}{2^{2k}} + 3uq_m^3\delta_2^2 \\ &+ \frac{3(n+1)^3 u}{2^{2n}} + \frac{2n^2 uq_m^2}{2^{2n}} + \frac{128n^3 uq_m^3}{2^{3n}}.\end{aligned}$$

Theorem 13 shows an improved result from the previous work [21]. We require the hash function to satisfy an additional assumption called δ-AU$^{(2)}$: For any $M_1 \neq M_1'$ and $M_2 \neq M_2'$ in $\mathcal{M}$, H is δ-AU$^{(2)}$ if:

$$\Pr_{K \xleftarrow{\$} \mathcal{K}}[\mathsf{H}_K(M_1) = \mathsf{H}_K(M_1') \wedge \mathsf{H}_K(M_2) = \mathsf{H}_K(M_2')] \leq \delta^2.$$

We remark that the cross-collision resistance between H^1, H^2 that are originally required in [21] automatically holds by the domain separation. We defer the proof to Section 7.2.

Theorem 13 (Proof is in Section 7.2). *Let n, k, u, p and q_m be positive integers. Let hash functions H^1, H^2 be δ-regular, δ-AU and δ-$AU^{(2)}$. Then, it holds that*

$$\mathsf{Adv}^{\mathsf{mu\text{-}prf}^*}_{\mathsf{DbHtS}}(u, q_m, p) \leq \frac{2upq_m\delta}{2^k} + \frac{2u^2q_m^2\delta}{2^k} + 10uq_m^2\delta^{\frac{3}{2}} + \frac{3upq_m}{2^{\frac{n}{2}+k}} + \frac{2u^2}{2^{2k}} + \frac{47uq_m^3\delta^{\frac{1}{4}}}{2^{2n}}.$$

7.1 Proof of Theorem 12

Transcript from the ideal and real world. We consider an arbitrary distinguisher $\mathcal{D}$ in the information-theoretic setting. Whenever the distinguisher makes a query, it obtains two types of information depending on the query, sometimes called entry, in both the ideal world and the real world:

- *Ideal-cipher queries*: For each primitive query on ideal cipher E with input x, we associate it with an entry $(\mathsf{prim}, J, x, y, +)$ for $J \in \mathcal{K}$ and $x, y \in \{0,1\}^n$. For each primitive query on the inverse of ideal cipher E^{-1} with input y, we associate it with an entry $(\mathsf{prim}, J, x, y, -)$ for $J \in \mathcal{K}$ and $x, y \in \{0,1\}^n$.
- *Construction queries:* For each construction query on DbHtS from user i with message M, we associate it with an entry (eval, i, M, T).

Let $(\mathsf{eval}, i, M_a^i, T_a^i)$ be the entry obtained when $\mathcal{D}$ makes the a-th query to user i. Let l_a^i be the number of blocks of M_a^i and let l be the maximal number of blocks among these uq_m construction queries. During the computation of $(\mathsf{eval}, i, M_a^i, T_a^i)$, let Σ_a^i, Ψ_a^i be the internal outputs of hash function H, namely $\Sigma_a^i = H^1_{K_{h,1}}(M_a^i)$ and $\Psi_a^i = H^2_{K_{h,2}}(M_a^i)$, respectively. Let U_a^i, Q_a^i be the outputs of ideal cipher E deployed in DbHtS with inputs Σ_a^i and Ψ_a^i, namely $U_a^i = E(K_i, 0\|\Sigma_a^i)$ and $Q_a^i = E(K_i, 1\|\Psi_a^i)$, respectively.

For a key $J \in \{0,1\}^k$, let $\mathbb{P}(J)$ be the set of entries $(\mathsf{prim}, J, x, y, *)$ associating with the primitive query on the ideal cipher E with key J. Let $\mathbb{Q}(J)$ be the set of entries $(\mathsf{eval}, i, M_a^i, T_a^i)$ associating with the construction query on DbHtS with the key such that $K_i = J$.

In the real world, after the distinguisher finishes all its queries, we will further give the following information to the distinguished:

1. the keys (K_h^i, K_i) for each user i, and
2. the internal values U_a^i, Q_a^i for each user i and its corresponding query a. In the ideal world, we will instead give the distinguisher $(K_h^i, K_i) \leftarrow_{\$} \{0,1\}^{2k} \times \{0,1\}^k$, independent of its queries.
3. In addition, we will give the distinguisher dummy values U_a^i and Q_a^i computed by the simulation oracle $\mathrm{SIM}(\mathbb{Q}(J))$ (the same as that in Fig.4 in [37]).

Both a transcript in the ideal world and the real world consists of

1. the revealed key pair (K_h^i, K_i) for each of u users,
2. the internal values U_a^i and Q_a^i for each of u users and each of their q_m construction queries, and

3. the p primitive queries and uq_m construction queries.

Bad Transcript Analysis and Interpretations. We now define bad transcripts and compute the probabilities that each bad event happens. Let T_{id} and T_{re} be random variables following the distribution of the transcripts in the real world and the ideal world, respectively. Let bad_i be the event that T_{id} satisfies the i-th bad event. We call the transcript *bad* if any bad events happen, and *good* otherwise. We refer [37, Section 3] for a more detailed description of most bad events and their analysis; the analysis for our cases is done analogously with small tweaks. The events [37, $\mathsf{bad}_{15}, \mathsf{bad}_{16}$] are excluded by the fine-tuned ideal world (bad_{15}) and the domain separation (bad_{16}) in the analysis below. Instead, we consider new bad_{15} below for

1. There exists user i such that $K_i = K^i_{h,1}$ or $K_i = K^i_{h,2}$. For each user, this event happens with probability at most $\frac{2}{2^k}$. Then, we have by the union bound that
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_1\right] \leq \frac{2u}{2^k}.$$
2. There exists user i such that both its key (K_i and ($K^i_{h,1}$ or $K^i_{h,2}$)) has been used by other user j or queried by primitive query $(\mathsf{prim}, J, x, y, *)$. Then, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_2\right] \leq \frac{u(3u+p)(6u+2p)}{2^{2k}}.$$

Note that $\neg\mathsf{bad}_1 \wedge \neg\mathsf{bad}_2$ guarantees that any user i has at least one fresh key. Further, excluding bad_3 defined below guarantees that the distinguisher cannot control the output of the hash function by issuing primitive queries that happen to use the same key and query a message block used in the hashing process.

3. There are two queries $(\mathsf{eval}, i, M^i_a, T^i_a)$ and $(\mathsf{prim}, J, x, y, +)$ of $\mathcal{D}$ such that the hash key used in the construction query is the same as the key used in the primitive query, namely $K^i_{h,1} = J$ or $K^i_{h,2} = J$; and one of the message block used in the construction query is queried by the primitive query, namely $x \in M'$, where M' might be a block in M^i_a or a proceed block during the hashing process depending on the construction of H. Then, if $p + uq_m l \leq 2^{n-2}$, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_3\right] \leq \frac{2upq_m l}{2^{k+n}}.$$

Excluding the following events, bad_4 and bad_5, guarantees that neither the inputs nor outputs of E_{K_i} collide with those in the primitive queries when $K_i = J$.

4. There are two queries $(\mathsf{eval}, i, M^i_a, T^i_a)$ and $(\mathsf{prim}, J, x, y, *)$ made by $\mathcal{D}$ such that $K_i = J$ and either $\Sigma^i_a = x$ or $\Psi^i_a = x$. Then, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_4 | \neg\mathsf{bad}_2\right] \leq \frac{2upq_m\delta_1}{2^k}.$$

5. There are entries $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{prim}, J, x, y, *)$ such that $K_i = J$ and either $U_a^i = y$ or $Q_a^i = y$. The event $U_a^i = y$ or $Q_a^i = y$ is the same as $\Phi_{K_i}(U_a^i) = y$ or $\Phi_{K_i}(Q_a^i) = y$ (Recall Φ is the partial function used to simulate a random permutation and defined in Fig.4 [37]). Then, if $p+uq_m l \leq 2^{n-2}$, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_5 | \neg\mathsf{bad}_2\right] \leq \frac{8upq_m}{2^{k+n}}.$$

Excluding the following two bad events bad_6 and bad_7 guarantees that if $K_i = K_j$ then all the inputs of E_{K_i} are distinct from those of E_{K_i}.

6. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that $K_i = K_j$ and $\Sigma_a^i = \Sigma_b^j$ for some entry $(\mathsf{eval}, j, M_b^j, T_b^j)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_6 | \neg\mathsf{bad}_2\right] \leq \frac{u^2 q_m^2 \delta_1}{2^k}.$$

7. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that $K_i = K_j$ and $\Psi_a^i = \Psi_b^j$ for some entry $(\mathsf{eval}, j, M_b^j, T_b^j)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_7 | \neg\mathsf{bad}_2\right] \leq \frac{u^2 q_m^2 \delta_1}{2^k}.$$

Excluding the following bad event bad_8 guarantees that if $K_i = K_{h,1}^j$ or $K_i = K_{h,2}^j$, then the inputs to E_{K_i} do not collide with the inputs to the hash part with key $K_{h,1}^j$ or $K_{h,2}^j$, respectively.

8. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that $K_i = K_{h,1}^j$ and $\Sigma_a^i = {M_b^j}'$, or $K_i = K_{h,2}^j$ and $\Psi_a^i = {M_b^j}'$ for some entry $(\mathsf{eval}, i, M_b^j, T_b^j)$, where ${M_b^j}'$ is either one of message blocks of M_b^j or a proceed block of M_b^j during the hashing process depending on the construction of H. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_8\right] \leq \frac{2u^2 q_m^2 l \delta_1}{2^k}.$$

Excluding the following bad event bad_9 guarantees that for every pair of construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ from each user, at least one of $0\|\Sigma_a^i$ and $1\|\Psi_a^i$ is fresh for the construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$. In other words, at least one of inputs of the ideal cipher E_{K_i} is fresh.

9. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that $\Sigma_a^i = \Sigma_b^i$ and $\Psi_a^i = \Psi_b^i$ for some entry $(\mathsf{eval}, i, M_b^i, T_b^i)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_9\right] \leq u q_m^2 \delta_2^2.$$

Excluding the following two bad events bad_{10} and bad_{11} guarantees that the partial function (Def. in Fig.4 [37]) behaves indistinguishably from a random permutation for every pair of construction queries.

10. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that either $\Sigma_a^i = \Sigma_b^i$ or $\Sigma_a^i = \Psi_b^i$, and either $U_a^i = U_b^i$ or $U_a^i = Q_b^i$ for some entry $(\mathsf{eval}, i, M_b^i, T_b^i)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{10}\right] \leq \frac{2uq_m^2(\delta_1 + \delta_2)}{2^n}.$$

11. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that either $\Psi_a^i = \Sigma_b^i$ or $\Psi_a^i = \Psi_b^i$, and either $Q_a^i = U_b^i$ or $Q_a^i = Q_b^i$ for some entry $(\mathsf{eval}, i, M_b^i, T_b^i)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{11}\right] \leq \frac{2uq_m^2(\delta_1 + \delta_2)}{2^n}.$$

Excluding the following bad event and event bad_9 guarantees that for every triple of construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$, $(\mathsf{eval}, i, M_b^i, T_b^i)$ and $(\mathsf{eval}, i, M_c^i, T_c^i)$ from each user, at least one of $0\|\Sigma_a^i$ and $1\|\Psi_a^i$ is fresh for the construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$. In other words, at least one of inputs of the ideal cipher E_{K_i} is fresh.

12. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that

$$\left[\Sigma_a^i = \Sigma_b^i \text{ and } \Psi_a^i = \Psi_c^i\right] \text{ or } \left[\Sigma_a^i = \Sigma_c^i \text{ and } \Psi_a^i = \Psi_b^i\right]$$

for some entries $(\mathsf{eval}, i, M_b^i, T_b^i)$ and $(\mathsf{eval}, i, M_c^i, T_c^i)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{12}\right] \leq 2uq_m^3\delta_2^2.$$

Excluding the following two bad events $\mathsf{bad}_{13}, \mathsf{bad}_{14}$ guarantees that the partial function (Def. in Fig.4 of [37]) behaves indistinguishably with a random permutation for every triple of construction query.

13. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that either $\Sigma_a^i = \Sigma_b^i$ or $\Sigma_a^i = \Psi_b^i$, and either $U_a^i = U_c^i$ or $U_a^i = Q_c^i$ for some entry $(\mathsf{eval}, i, M_b^i, T_b^i)$ and $(\mathsf{eval}, i, M_c^i, T_c^i)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{13}\right] \leq \frac{2uq_m^3(\delta_1 + \delta_2)}{2^n}.$$

14. There is a construction entry $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that either $\Psi_a^i = \Sigma_b^i$ or $\Psi_a^i = \Psi_b^i$, and either $Q_a^i = U_c^i$ or $Q_a^i = Q_c^i$ for some entry $(\mathsf{eval}, i, M_b^i, T_b^i)$ and $(\mathsf{eval}, i, M_c^i, T_c^i)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{14}\right] \leq \frac{2uq_m^3(\delta_1 + \delta_2)}{2^n}.$$

Excluding the following bad event guarantees that there are no more than n users who share the same ideal cipher key.

15. There are no $i_1, \ldots, i_n \in [u]$ such that $K_{i_1} = \cdots = K_{i_n}$. Then we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{15}\right] = \frac{\binom{u}{n}}{2^{k(n-1)}} \leq \frac{u^2}{2^{k+n}}.$$

The overall probability of the bad event is bounded as follows:

$$\begin{aligned}\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}\right] \leq \frac{2u}{2^k} &+ \frac{u(3u+p)(6u+2p)}{2^{2k}} + \frac{2upq_m l}{2^{k+n}} + \frac{2upq_m\delta_1}{2^k} \\ &+ \frac{8upq_m}{2^{k+n}} + \frac{4u^2q_m^2 l\delta_1}{2^k} + 3uq_m^3\delta_2^2 + \frac{8uq_m^3(\delta_1+\delta_2)}{2^n} + \frac{u^2}{2^{k+n}}. \quad (40)\end{aligned}$$

Good Transcript Analysis. The following analysis is built upon the analysis presented in [37], highlighting only the modifications relevant to the context of our paper. Let τ denote a good transcript and $S(J), F(J), Q(J)$ are defined in [37, Figure 4] and g is defined in [37, page 15].

We first compute

$$\Pr\left[\mathsf{T}_{\mathsf{id}} = \tau\right] = \frac{1}{2^{2uk}(N-1)^{uq_m}} \prod_{J \in \{0,1\}^k} \left(\frac{1}{|S(J)|} \cdot \frac{1}{(N-2\,|F(J)|)_g} \prod_{i=0}^{|\mathbb{P}(J)|-1} \frac{1}{N-2\,|F(J)|-g-i} \right),$$

and

$$\Pr\left[\mathsf{T}_{\mathsf{re}} = \tau\right] = \frac{1}{2^{2uk}} \prod_{J \in \{0,1\}^k} \left(\frac{1}{(N)_{|Q(J)|+|F(J)|+g}} \prod_{i=0}^{|\mathbb{P}(J)|-1} \frac{1}{N-|Q(J)|-|F(J)|-g-i} \right).$$

Then, we have

$$
\begin{aligned}
\frac{\Pr[\mathsf{T}_{\mathsf{re}}=\tau]}{\Pr[\mathsf{T}_{\mathsf{id}}=\tau]} &\geq (N-1)^{uq_m} \prod_{J\in\{0,1\}^k} \frac{|S(J)|\,(N-2\,|F(J)|)_g}{(N)_{|Q(J)|+|F(J)|+g}} \\
&\geq \prod_{J\in\{0,1\}^k} \frac{(N-1)^{|Q(J)|}(N-2\,|F(J)|)_g}{(N)_{|Q(J)|+|F(J)|+g}} \cdot |S(J)| \\
&\geq \prod_{J\in\{0,1\}^k} \frac{(N-1)^{|Q(J)|}(N-2\,|F(J)|)_g}{(N)_{|Q(J)|+|F(J)|+g}} \frac{(N)_{2|F(J)|}}{(N-1)^{|F(J)|}} \\
&\quad \times \left(1-\frac{2\,|F(J)|^2}{N^2}-\frac{128\,|F(J)|^3}{N^3}-\frac{8(n+1)^3}{3N^2}\right) \quad \text{(by Theorem 6)} \\
&\geq \prod_{J\in\{0,1\}^k} \frac{(N-1)^{|Q(J)|-|F(J)|}}{(N-2\,|F(J)|-g)_{|Q(J)|-|F(J)|}} \\
&\quad \times \left(1-\frac{2\,|F(J)|^2}{N^2}-\frac{128\,|F(J)|^3}{N^3}-\frac{8(n+1)^3}{3N^2}\right) \\
&\geq \prod_{J\in\{0,1\}^k} \frac{(N-1)^{|Q(J)|-|F(J)|}}{(N-2\,|F(J)|-g)_{|Q(J)|-|F(J)|}} \\
&\quad \times \left(1-\frac{2n^2q_m^2}{N^2}-\frac{128n^3q_m^3}{N^3}-\frac{8(n+1)^3}{3N^2}\right) \\
&\qquad\qquad (\because |F(J)|\leq nq_m, \text{ guaranteed by } \neg\mathsf{bad}_{15}) \\
&\geq 1-\frac{2un^2q_m^2}{N^2}-\frac{128un^3q_m^3}{N^3}-\frac{8(n+1)^3u}{3N^2}, \qquad (41)
\end{aligned}
$$

where the last line is because there are at most u used J, so at most u product terms.

Conclude the Proof. From Equations (40) and (41), define

$$
\epsilon_1 \stackrel{\text{def}}{=} \frac{2n^2uq_m^2}{2^{2n}}+\frac{128n^3uq_m^3}{2^{3n}}+\frac{8(n+1)^3u}{3\cdot 2^{2n}}
$$

and

$$
\begin{aligned}
\epsilon_2 \stackrel{\text{def}}{=} \frac{2u}{2^k}+\frac{2upq_m\delta_1}{2^k}+\frac{4u^2q_m^2l\delta_1}{2^k}+\frac{8uq_m^3(\delta_1+\delta_2)}{2^n}+\frac{2upq_ml}{2^{k+n}} \\
+\frac{8upq_m}{2^{k+n}}+\frac{u^2}{2^{k+n}}+\frac{u(3u+p)(6u+2p)}{2^{2k}}+3uq_m^3\delta_2^2.
\end{aligned}
$$

Then by Lemma 1, we conclude that

$$\begin{aligned}\mathsf{Adv}^{\text{mu-prf}^*}_{\mathsf{DbHtS}}(u, q_m, p) \leq{} & \frac{2u}{2^k} + \frac{2upq_m\delta_1}{2^k} + \frac{4u^2q_m^2 l\delta_1}{2^k} + \frac{8uq_m^3(\delta_1+\delta_2)}{2^n} + \frac{2upq_m l}{2^{k+n}} \\ & + \frac{8upq_m}{2^{k+n}} + \frac{u^2}{2^{k+n}} + \frac{u(3u+p)(6u+2p)}{2^{2k}} + 3uq_m^3\delta_2^2 + \frac{3(n+1)^3u}{2^{2n}} \\ & + \frac{2n^2uq_m^2}{2^{2n}} + \frac{128n^3uq_m^3}{2^{3n}}.\end{aligned}$$

7.2 Proof of Theorem 13

The general idea of the proof follows the proof of [21, Theorem 1]. The concrete proof diverges in threefold. First, we analyze the multi-user security of DbHtS in the fine-tuned ideal world, which excludes the bad event Bad-Tag unavoidable in the analysis of [21]. Second, our DbHtS construction also assumes the hash function is δ-AU$^{(2)}$. Third, we explicitly introduce q_m instead of approximating it as q in the analysis. These variances not only modify the calculations associated with certain bad events presented in [21] but also lead to consequential shifts in the final statement (Theorem 13). For the sake of completeness, we give the full proof in the following. Note that if $q_m^2\delta^{\frac{3}{2}} \geq 1$, then Theorem 13 trivially holds. Hence in the following, we prove Theorem 13 for the case of $q_m^2\delta^{\frac{3}{2}} < 1$.

Transcript From the Ideal and Real World. We consider an arbitrary distinguisher $\mathcal{D}$ in the information-theoretic setting. After the distinguisher finishes querying, it obtains two types of information in both of the ideal world and the real world

- *Ideal-cipher queries:* for each primitive query on ideal cipher E with input x, we associate it with an entry $(\mathsf{prim}, J, x, y, +)$ for $J \in \mathcal{K}$ and $x, y \in \{0,1\}^n$. For each primitive query on the inverse of ideal cipher E^{-1} with input y, we associate it with an entry $(\mathsf{prim}, J, x, y, -)$ for $J \in \mathcal{K}$ and $x, y \in \{0,1\}^n$.
- *Construction queries:* for each construction query on DbHtS from user i with message M, we associate it with an entry (eval, i, M, T).

Let $(\mathsf{eval}, i, M_a^i, T_a^i)$ be the entry obtained when $\mathcal{D}$ makes the a-th query to user i. During the computation of $(\mathsf{eval}, i, M_a^i, T_a^i)$, let Σ_a^i, Ψ_a^i be the internal outputs of hash function H, namely $\Sigma_a^i = H^1_{K_{h,1}}(M_a^i)$ and $\Psi_a^i = H^2_{K_{h,2}}(M_a^i)$, respectively. Let U_a^i, Q_a^i be the outputs of ideal cipher E deployed in DbHtS with inputs Σ_a^i and Ψ_a^i, namely $U_a^i = E(K_i, 0\|\Sigma_a^i)$ and $Q_a^i = E(K_i, 1\|\Psi_a^i)$, respectively. To make it a bit easy to read, we use the term "block cipher key" to refer to the key K_i for user i and "ideal cipher key" to refer to the key J used in an Ideal-cipher (primitive) query. Let s denote the total number of distinct ideal cipher key used during the evaluation of primitive queries. Let r denote the total number of distinct block cipher keys that collide with ideal cipher key used during the evaluation of primitive queries.

In the real world, after the distinguisher finishes all its queries, we will further give it: 1) the keys (K_h^i, K_i) for each user i and 2) the internal values

$(\Sigma^i_a, \Psi^i_a, U^i_a, Q^i_a)$ for each user i and its corresponding query a. In the ideal world, we will instead give the distinguisher $(K^i_h, K_i) \leftarrow_\$ \{0,1\}^{2k} \times \{0,1\}^k$, independent of its queries. In addition, we will give the distinguisher dummy values $(\Sigma^i_a, \Psi^i_a, U^i_a, Q^i_a)$. All these values are computed by the simulation oracle shown in Algorithm 3. Note that we remove $\mathsf{Bad\text{-}Tag}$ event since we are assuming the fine-tuned ideal world while there is an additional bad event, dubbed $\mathsf{Bad5}$, to achieve better security.

Both a transcript in the ideal world and the real world consists of

1. the revealed keys (K^i_h, K_i) for each of u users,
2. the internal values $(\Sigma^i_a, \Psi^i_a, U^i_a, Q^i_a)$ for each of u users and each of their q_m construction queries,
3. and the p primitive queries and uq_m construction queries.

Algorithm 3 Offline oracle in the ideal world

$(K^i_{h,1}, K^i_{h,2})_{i\in[u]} \leftarrow_\$ \mathcal{K}_h \times \mathcal{K}_h$
$(K_i)_{i\in[u]} \leftarrow_\$ \{0,1\}^k$
$(\Sigma^i_a, \Psi^i_a)_{(i,a)\in[u]\times[q_m]} \leftarrow (H^1_{K_{h,1}}(M^i_a), H^2_{K_{h,2}}(M^i_a))_{(i,a)\in[u]\times[q_m]}$
if $\boxed{\mathsf{BadK}=1 \vee \mathsf{Bad1}=1 \vee \mathsf{Bad2}=1 \vee \mathsf{Bad3}=1 \vee \mathsf{Bad4}=1 \vee \mathsf{Bad5}=1}$ **then** aborts
$\mathbb{Q}^= \stackrel{\text{def}}{=} \{(i,a) \in [u]\times[q_m] : \exists(\mathsf{prim}, K_i, x, y, *); \forall(\mathsf{prim}, K_i, x, y, *), x \neq \Sigma^i_a, x \neq \Psi^i_a\}$
$\mathbb{I}^= \stackrel{\text{def}}{=} \{i \in [u] : (i,*) \in \mathbb{Q}^=\} = \mathbb{I}^=_{i_1} \sqcup \cdots \sqcup \mathbb{I}^=_{i_r}$ $\triangleright$ $i \in \mathbb{I}^=_{i_j}$ if K_{i_j} is used in primitive queries, where $i_j \in [s]$ as there are s distinct ideal-cipher key
for $j \leftarrow 1$ to r **do**
  $\forall i \in \mathbb{I}^=_{i_j}$ let Σ^i_a be not fresh in $(\Sigma^i_1, \ldots, \Sigma^i_{q_m})$ for some construction query $(\mathsf{eval}, i, M^i_a, T^i_a)$
  Let $\mathsf{Dom}(K_{i_j}) \stackrel{\text{def}}{=} \{x : (\mathsf{prim}, K_{i_j}, x, y, *)\}$ and $\mathsf{Ran}(K_{i_j}) \stackrel{\text{def}}{=} \{y : (\mathsf{prim}, K_{i_j}, x, y, *)\}$
  if $0\|\Sigma^i_a \notin \mathsf{Dom}(K_{i_j})$ **then** $\mathrm{P}_{i_j}(\Sigma^i_a) \leftarrow U^i_a \leftarrow_\$ \{0,1\}^n \setminus \mathsf{Ran}(K_{i_j}), Q^i_a \leftarrow U^i_a \oplus T^i_a$
  else $U^i_a \leftarrow \mathrm{P}_{i_j}(\Sigma^i_a), Q^i_a \leftarrow U^i_a \oplus T^i_a$
  if $Q^i_a \in \mathsf{Ran}(K_{i_j})$ **then** $\boxed{\mathsf{Bad\text{-}Samp} \leftarrow 1}$, aborts
  else $\mathsf{Dom}(K_{i_j}) \leftarrow \mathsf{Dom}(K_{i_j}) \bigcup \{0\|\Sigma^i_a, 1\|\Psi^i_a\}, \mathsf{Ran}(K_{i_j}) \leftarrow \mathsf{Ran}(K_{i_j}) \bigcup \{U^i_a, Q^i_a\}$
$\mathbb{Q}^{\neq} \stackrel{\text{def}}{=} \{(i,a) \in [u]\times[q_m] : \forall(\mathsf{prim}, J, x, y, *), J \neq K_i\}$
$\mathbb{I}^{\neq} \stackrel{\text{def}}{=} \{i \in [u] : (i,*) \in \mathbb{Q}^{\neq}\} = \mathbb{I}^{\neq}_{i_1} \sqcup \cdots \sqcup \mathbb{I}^{=}_{i'_r}$ $\triangleright$ $i, j \in \mathbb{I}^{\neq}_{i_j}$ if $K_i = K_j$
$\forall j \in [r'] : \widetilde{\Sigma^{i_j}} = \bigcup_{i\in\mathbb{I}^{\neq}_{i_j}} \{\Sigma^i_1, \ldots, \Sigma^i_{q_m}\}, \widetilde{\Psi^{i_j}} = \bigcup_{i\in\mathbb{I}^{\neq}_{i_j}} \{\Psi^i_1, \ldots, \Psi^i_{q_m}\}$
$\forall j \in [r'] : (U^i_a, Q^i_a)_{i\in\mathbb{I}^{\neq}_{i_j}, a\in[q_m]} \leftarrow_\$ \mathcal{S}_{i_j}$ where $\mathcal{S}_{i_j} \stackrel{\text{def}}{=} \{\bigcup_{i\in\mathbb{I}^{\neq}_{i_j}, a\in[q_m]} \{Z^i_{a,1}, Z^i_{a,2}\} \in (\{0,1\}^n)^{|\widetilde{\Sigma^{i_j}}|+|\widetilde{\Psi^{i_j}}|} : Z^i_{a,1} \oplus Z^i_{a,2} = T^i_a\}$
  return $(\Sigma^i_a, \Psi^i_a, U^i_a, Q^i_a)_{(i,a)\in[u]\times[q_m]}$

Bad Transcript Analysis and Interpretations. We now give the definition of bad transcripts and compute their corresponding probabilities. Let T_{id}

and T_{re} be random variables following the distribution of the transcripts in the real world and the ideal world, respectively. We define and calculate the following bad events that are with non-zero probability in T_{id}.

BadK**:** There exists distinct user i_1 and i_2 such that both block-cipher key and (one of) hash key collide, i.e., $K_{i_1} = K_{i_2} \wedge (K_{h,1}^{i_1} = K_{h,1}^{i_2} \vee K_{h,2}^{i_1} = K_{h,2}^{i_2})$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{BadK}\right] \leq \frac{2u^2}{2^{2k}}.$$

$\mathsf{Bad1}$**:** There exists a construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ such that its corresponding block-cipher key and (one of) hash output collide with a primitive query $(\mathsf{prim}, J, x, y, *)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{Bad1}\right] \leq \frac{2upq_m\delta}{2^k}.$$

$\mathsf{Bad2}$**:** Either $\mathsf{B.21}$ or $\mathsf{B.22}$ happens, so we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{Bad2}|\neg\mathsf{BadK}\right] \leq \frac{uq_m^2\delta}{2^n} + \frac{2u^2q_m^2\delta}{2^k}.$$

$\mathsf{B.21}$**:** There exists a user i whose two construction queries both collide on tag and (one of) hash output. That is, there exists $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ such that $T_a^i = T_b^i \wedge (\Sigma_a^i = \Sigma_b^i \vee \Psi_a^i = \Psi_b^i)$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.21}|\neg\mathsf{BadK}\right] \leq \frac{uq_m^2\delta}{2^n}.$$

$\mathsf{B.22}$**:** There exists users i_1 and i_2 whose construction queries both collide on tag and (one of) hash output. That is, there exists $(\mathsf{eval}, i_1, M_a^{i_1}, T_a^{i_1})$ and $(\mathsf{eval}, i_2, M_b^{i_2}, T_b^{i_2})$ such that $T_a^{i_1} = T_b^{i_2} \wedge (\Sigma_a^{i_1} = \Sigma_b^{i_2} \vee \Psi_a^{i_1} = \Psi_b^{i_2})$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.22}|\neg\mathsf{BadK}\right] \leq \frac{2u^2q_m^2\delta}{2^k}.$$

$\mathsf{Bad3}$**:** Either $\mathsf{B.31}$ or $\mathsf{B.32}$ happens, so we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{Bad3}\right] \leq 2uq_m^2\delta^{\frac{3}{2}}.$$

$\mathsf{B.31}$: There exists user i such that $\left|\{(a,b) : a < b \wedge \Sigma_a^i = \Sigma_b^i\}\right| \geq L_1$. Let $L_1 = \frac{q_m^2\delta}{2} + \frac{q_m\delta^{\frac{1}{4}}}{2}$. Let $\mathbb{I}_{a,b}^i$ be the indicator random variable which takes the value 1 if $\Sigma_a^i = \Sigma_b^i$; 0 otherwise. Let $\mathbb{I}^i = \sum_{a<b} \mathbb{I}_{a,b}^i$.
We calculate

$$\mathbf{Ex}\left[\mathbb{I}^i\right] = \sum_{a<b} \Pr\left[\Sigma_a^i = \Sigma_b^i\right] \leq \frac{q_m^2\delta}{2},$$

and, by assuming the hash function is δ-AU$^{(2)}$,

$$\mathbf{Var}\left[\mathbb{I}^i\right] \leq \mathbf{Ex}\left[(\mathbb{I}^i)^2\right] = \mathbf{Ex}\left[\left(\sum_{a<b} \mathbb{I}^i_{a,b}\right)\left(\sum_{c<d} \mathbb{I}^i_{c,d}\right)\right] \leq \frac{q_m^4 \delta^2}{4}.$$

Then, by Lemma 3, we have

$$\begin{aligned}
\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.31}\right] &\leq \sum_{i\in[u]} \Pr\left[\mathbb{I}^i \geq L_1\right] \\
&\leq \frac{u q_m^4 \delta^2}{(2L_1 - q_m^2 \delta)^2} \\
&= \frac{u q_m^4 \delta^2}{q_m^2 \delta^{\frac{1}{2}}} && (\because \text{Plug in } L_1) \\
&\leq u q_m^2 \delta^{\frac{3}{2}}.
\end{aligned}$$

B.32: There exists user i such that $\left|\{(a,b) : a < b \wedge \Psi^i_a = \Psi^i_b\}\right| \geq L_1$. Similarly, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.32}\right] \leq u q_m^2 \delta^{\frac{3}{2}}.$$

Bad4**:** One of B.41, B.42, and B.43 happens, so we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{Bad4}\right] \leq \frac{u q_m^2 \delta^2}{2} + 4 u q_m^2 \delta^{\frac{3}{2}} \leq \frac{9 u q_m^2 \delta^{\frac{3}{2}}}{2}.$$

B.41: There exists a user i whose two construction queries both two hash outputs collide. That is, there exists $(\mathsf{eval}, i, M^i_a, T^i_a)$ and $(\mathsf{eval}, i, M^i_b, T^i_b)$ such that $\Sigma^i_a = \Sigma^i_b \wedge \Psi^i_a = \Psi^i_b$. Then, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.41}\right] \leq \frac{u q_m^2 \delta^2}{2}.$$

B.42: There exists a user i and its tuple of four construction queries indices a, b, c, d such that $\Sigma^i_a = \Sigma^i_b \wedge \Psi^i_b = \Psi^i_c \wedge \Sigma^i_c = \Sigma^i_d$. Then, we have

$$\begin{aligned}
\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.42} | \neg \mathsf{B.31}\right] &\leq \sum_{i\in[u]} L_1^2 \delta \\
&= \left(\frac{q_m^2 \delta}{2} + \frac{q_m \delta^{\frac{1}{4}}}{2}\right)^2 \delta && (\because \text{Plug in } L_1) \\
&\leq \frac{1}{2} q_m^4 \delta^3 + \frac{1}{2} q_m^2 \delta^{\frac{3}{2}} && (\text{by Lemma } 6) \\
&\leq q_m^2 \delta^{\frac{3}{2}}. && (\because q_m^2 \delta^{\frac{3}{2}} < 1)
\end{aligned}$$

Further,

$$\begin{aligned}
\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.42}\right] \leq\ & \Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.42} | \neg \mathsf{B.31}\right] + \Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.31}\right] \\
\leq\ & 2 q_m^2 \delta^{\frac{3}{2}}.
\end{aligned}$$

Note that the summation of $\Pr[\mathsf{T_{id}} \in \mathsf{B.31}]$ will happen again when we compute the total probability of bad events, so this is indeed redundant computation, but we leave it for simplicity of the proof.

B.43: There exists a user i and its tuple of four construction queries indices a, b, c, d such that $\Psi_a^i = \Psi_b^i \wedge \Sigma_b^i = \Sigma_c^i \wedge \Psi_c^i = \Psi_d^i$. Similarly, we have

$$\Pr[\mathsf{T_{id}} \in \mathsf{B.43}] \leq 2q_m^2 \delta^{\frac{3}{2}}.$$

Bad5: There exists user i such that

$$\left|\{(a,b,c) \in [q_m]^{*3} : \Sigma_a^i = \Sigma_b^i \wedge \Psi_b^i = \Psi_c^i\}\right| \geq L_2.$$

Let $\mathbb{I}_{a,b,c}^i$ be the indicator random variable which takes the value 1 if $\Sigma_a^i = \Sigma_b^i \wedge \Psi_b^i = \Psi_c^i$; 0 otherwise. Let

$$\mathbb{I}^i = \sum_{(a,b,c) \in [q_m]^{*3}} \mathbb{I}_{a,b,c}^i$$

and $L_2 = q_m^{\frac{1}{3}} \delta^{\frac{1}{2}} 2^{\frac{2}{3}n}$. Since we assume δ-AU$^{(2)}$,

$$\mathbf{Ex}\left[\mathbb{I}^i\right] = \sum_{(a,b,c) \in [q_m]^{*3}} \Pr\left[\Sigma_a^i = \Sigma_b^i \wedge \Psi_b^i = \Psi_c^i\right] \leq q_m^3 \delta^2,$$

Then, by Lemma 2, we have

$$\begin{aligned}
\Pr[\mathsf{T_{id}} \in \mathsf{Bad5}] &\leq \sum_{i \in [u]} \Pr\left[\mathbb{I}^i \geq L_2\right] \\
&\leq \frac{u q_m^3 \delta^2}{L_2} \\
&= \frac{u q_m^{\frac{8}{3}} \delta^{\frac{3}{2}}}{2^{\frac{2}{3}n}}. \qquad (\because \text{Plug in } L_2)
\end{aligned}$$

Bad-Samp: Bad-Samp happens if in the simulation oracle [line 12, Algorithm 3], the simulated output Q_a^i collides with any primitive query $(\mathsf{prim}, K_{i_j}, x, y, *)$ or previous simulated output. We bound Bad-Samp by the union of events BS1 and BS2. We assume $q = uq_m$.

BS1: There exists a primitive query $(\mathsf{prim}, J, x, y, *)$ and a user $i \in \mathbb{I}_{i_j}^{=}$ for some $j \in [r]$ such that both its output y collides with Q_a^i and its ideal-cipher key J equals to the block-cipher key K_{i_j}. Then we have we have

$$\Pr[\mathsf{T_{id}} \in \mathsf{BS1}] \leq \frac{2pq}{2^{n+k}} = \frac{2upq_m}{2^{n+k}}.$$

BS2: There exists a primitive query $(\mathsf{prim}, J, x, y, *)$ and two users $i, i' \in \mathbb{I}_{i_j}^{=}$ for some $j \in [r]$ such that both Q_a^i collides with $Q_a^{i'}$ and its ideal-cipher key J equals to the block-cipher key K_{i_j}. For BS2, we have

$$\Pr[\mathsf{T_{id}} \in \mathsf{BS2}] \leq \sum_{i=1}^{u} \frac{pq_m^2}{2^{n+k}} = \frac{upq_m^2}{2^{n+k}} = \frac{pqq_m}{2^{n+k}}.$$

We consider two cases. If $q \le 2^{n/2}$, then we have

$$\frac{pqq_m}{2^{n+k}} \le \frac{pq_m}{2^{\frac{n}{2}+k}} \le \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

On the other hand, if $q > 2^{n/2}$, we start with considering the first $\frac{q}{2^{\frac{n}{2}}}$ users. Similarly to [21, Inequality (25)], we define an event Aux as follows: if the key for any of first $\frac{q}{2^{\frac{n}{2}}}$ users collide with a primitive query key, we call Aux occurs. We can see

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{Aux}\right] \le \frac{\left(\frac{q}{2^{\frac{n}{2}}}\right)p}{2^k} \le \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

If $u \le \frac{q}{2^{\frac{n}{2}}}$, it is an upper bound of $\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \le \frac{upq_m}{2^{\frac{n}{2}+k}}$. Otherwise, $q = uq_m \ge \frac{q}{2^{\frac{n}{2}}} q_m$, which says $q_m \le 2^{n/2}$. Then we have

$$\frac{pqq_m}{2^{n+k}} \le \frac{pq}{2^{\frac{n}{2}+k}} = \frac{upq_m}{2^{\frac{n}{2}+k}},$$

and

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \le \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

To conclude, we have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{Bad\text{-}Samp}\right] \le \frac{2upq_m}{2^{n+k}} + \frac{upq_m}{2^{\frac{n}{2}+k}} \le \frac{3upq_m}{2^{\frac{n}{2}+k}}.$$

Summing up the probability of the bad events and define $\mathsf{bad} = \mathsf{BadK} \vee \mathsf{Bad1} \vee \mathsf{Bad2} \vee \mathsf{Bad3} \vee \mathsf{Bad4} \vee \mathsf{Bad5} \vee \mathsf{Bad\text{-}Samp}$, we have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{bad}\right] \le \frac{2u^2}{2^{2k}} + \frac{2upq_m\delta}{2^k} + \frac{2u^2q_m^2\delta}{2^k} + 8uq_m^2\delta^{\frac{3}{2}} + \frac{3upq_m}{2^{\frac{n}{2}+k}}. \tag{42}$$

Good Transcript Analysis. The following analysis aims to compute a lower bound for the ratio $\frac{\Pr[\mathsf{T_{re}}=\tau]}{\Pr[\mathsf{T_{id}}=\tau]}$ on a good transcript. We first consider the transcript for the construction query indexed by $\mathbb{Q}^{=}$. Recall

$$\mathbb{Q}^{=} \stackrel{\text{def}}{=} \left\{(i,a) \in [u] \times [q_m] : \exists(\mathsf{prim}, K_i, x, y, *); \forall(\mathsf{prim}, K_i, x, y, *), x \ne \Sigma_a^i, x \ne \Psi_a^i\right\}$$

as defined in Algorithm 3. For each $j \in [r]$ and each $i \in \mathbb{I}_{i_j}^{=}$, we consider the internal value sequence

$$(U_1^i, \ldots, U_{q_m}^i), (Q_1^i, \ldots, Q_{q_m}^i).$$

From this sequence, we construct a bipartite graph G_i, where the nodes in one partition represent values U_a^i and the nodes in the other represent Q_a^i. We connect the node representing U_a^i and Q_a^i with an edge labeled with T_a^i, where $U_a^i \oplus Q_a^i = T_a^i$. If $U_a^i = U_b^i$ where $a \ne b$, then we merge the corresponding nodes into a single one. We do the same thing if $Q_a^i = Q_b^i$ where $a \ne b$.

Since the transcript is good, we know that each component of G_i is acyclic, which is guaranteed by $\neg\mathsf{B.41}$. Guaranteed by $\neg\mathsf{B.42} \wedge \neg\mathsf{B.43}$, each component contains a path of length at most 3. Also, guaranteed by $\neg\mathsf{B.31} \wedge \neg\mathsf{B.32}$, the size of each component is restricted up to $L_1 = \frac{q_m^2 \delta}{2} + \frac{q_m \delta^{\frac{1}{4}}}{2}$. Furthermore, guaranteed by $\neg\mathsf{Bad1}$, the value of each vertex of the graph G_i is distinct from the input of any primitive query. Guaranteed by $\neg\mathsf{B.21}$, if two nodes are connected in G_i the label of their path cannot be zero. Guaranteed by $\neg\mathsf{B.22}$, if two distinct users i_1, i_2 whose keys collide, then their corresponding graph G_{i_1} and G_{i_2} are distinct. We use v_i to denote the size of the graph G_i, and w_i to denote the number of components of G_i.

We then consider the transcript for the construction query indexed by $\mathbb{Q}^{\neq}$. Recall $\mathbb{I}^{\neq} \stackrel{\text{def}}{=} \{i \in [u] : (i, *) \in \mathbb{Q}^{\neq}\} = \mathbb{I}^{\neq}_{i_1} \sqcup \cdots \sqcup \mathbb{I}^{=}_{i'_r}$ as defined in Algorithm 3. For each $j \in [r']$ and each $i \in \mathbb{I}^{=}_{i_j}$, we consider the internal value sequence

$$(U_1^i, \ldots, U_{q_m}^i), (Q_1^i, \ldots, Q_{q_m}^i).$$

Similarly, we can construct a bipartite graph H_i. We use v'_i to denote the size of the graph H_i, and w'_i to denote the number of components of H_i.

We now are ready to compute $\Pr[\mathsf{T}_{\mathsf{re}} = \tau]$, the probability of real-world hits a good transcript τ. Let p_j be the number of primitive query use the j-th ideal-cipher key. We have

$$\Pr[\mathsf{T}_{\mathsf{re}} = \tau] = \prod_{i=1}^{u} \frac{1}{2^{3k}} \cdot \left(\prod_{j=1}^{r} \frac{1}{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}^{=}_{i_j}} v_i\right)}} \right) \prod_{j \in [s] \setminus \{i_1, \ldots, i_r\}} \frac{1}{(2^n)_{p_j}} \left(\prod_{j=1}^{r'} \frac{1}{(2^n)_{\left(\sum_{i \in \mathbb{I}^{\neq}_{i_j}} v'_i\right)}} \right).$$

And for $\Pr[\mathsf{T}_{\mathsf{id}} = \tau]$, we have

$$\Pr[\mathsf{T}_{\mathsf{id}} = \tau] = \frac{1}{2^{nuq_m}} \prod_{i=1}^{u} \frac{1}{2^{3k}} \cdot \left(\prod_{j=1}^{r} \frac{1}{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}^{=}_{i_j}} w_i\right)}} \right) \prod_{j \in [s] \setminus \{i_1, \ldots, i_r\}} \frac{1}{(2^n)_{p_j}} \left(\prod_{j=1}^{r'} \frac{1}{\left|\mathcal{S}_{i_j}\right|} \right).$$

Plugging in the above two expressions, we have

$$\frac{\Pr\left[\mathsf{T}_{\mathsf{re}} = \tau\right]}{\Pr\left[\mathsf{T}_{\mathsf{id}} = \tau\right]} = 2^{nuq_m} \left(\prod_{j=1}^{r} \frac{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}_{i_j}^{=}} w_i\right)}}{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}_{i_j}^{=}} v_i\right)}} \right) \cdot \left(\prod_{j=1}^{r'} \frac{\left|\mathcal{S}_{i_j}\right|}{(2^n)_{\left(\sum_{i \in \mathbb{I}_{i_j}^{\neq}} v'_i\right)}} \right)$$

$$\geq 2^{nuq_m} \left(\prod_{j=1}^{r} \frac{1}{\left(2^n - p_j - \sum_{i \in \mathbb{I}_{i_j}^{=}} w_i\right)_{\left(\sum_{i \in \mathbb{I}_{i_j}^{=}} (v_i - w_i)\right)}} \right) \cdot \left(\prod_{j=1}^{r'} \frac{\left(1 - \delta_{i_j}\right) \cdot (2^n)_{\left(\sum_{i \in \mathbb{I}_{i_j}^{\neq}} v'_i\right)}}{2^{\left(n \sum_{i \in \mathbb{I}_{i_j}^{\neq}} (v'_i - w'_i)\right)} \cdot (2^n)_{\left(\sum_{i \in \mathbb{I}_{i_j}^{\neq}} v'_i\right)}} \right),$$

(∵ Plug in Theorem 5)

where

$$\delta_{i_j} \stackrel{\text{def}}{=} \sum_{i \in \mathbb{I}_{i_j}^{\neq}} \frac{9q_{c,i}^2 \sum_{1 \leq k \leq \alpha_i} c_k^2}{8 \cdot 2^{2n}} + \frac{31 q_{c,i} q_i^2}{2^{2n}} + \frac{16 q_i^4}{2^{3n}}$$

$$\leq \frac{9q_{c,i}^2 L_2^2}{8 \cdot 2^{2n}} + \frac{31 q_{c,i} q_m^2}{2^{2n}} + \frac{16 q_m^4}{2^{3n}}. \qquad (\because \neg\mathsf{Bad5})$$

We here explain more on the definition of δ_{i_j}. $i \in \mathbb{I}_{i_j}^{\neq}$ is the user index and H_i is the bipartite graph constructed mentioned above. We use q_i to denote the total number of edges in H_i and $q_{c,i}$ to denote the total number of edges of components in H_i with size larger than 2. We use α_i to denote the total number of components in H_i with size larger than 2 and c_k to denote the size of k-th component in the graph H_i.

We further lower bound the expression of $\frac{\Pr[\mathsf{T}_{\mathsf{re}}=\tau]}{\Pr[\mathsf{T}_{\mathsf{id}}=\tau]}$ in the following

$$
\begin{aligned}
\frac{\Pr\left[\mathsf{T}_{\mathsf{re}}=\tau\right]}{\Pr\left[\mathsf{T}_{\mathsf{id}}=\tau\right]} \geq & \left(\prod_{j=1}^{r} \frac{2^{nq_m|\mathbb{I}^{=}_{i_j}|}}{\left(2^n - p_j - \sum_{i\in\mathbb{I}^{=}_{i_j}} w_i\right)^{\left(\sum_{i\in\mathbb{I}^{=}_{i_j}} (v_i - w_i)\right)}}\right) \cdot \left(\prod_{j=1}^{r'} \frac{2^{nq_m|\mathbb{I}^{\neq}_{i_j}|}\cdot(1-\delta_{i_j})}{2^{\left(n\sum_{i\in\mathbb{I}^{\neq}_{i_j}} (v'_i - w'_i)\right)}}\right) \\
\geq & 1 - \sum_{j=1}^{r'} \delta_{i_j} \\
\geq & 1 - \sum_{j=1}^{r'} \sum_{i\in\mathbb{I}^{\neq}_{i_j}} \left(\frac{9q_{c,i}^2 L_2^2}{8\cdot 2^{2n}} + \frac{31 q_{c,i} q_m^2}{2^{2n}} + \frac{16 q_m^4}{2^{3n}}\right) \\
\geq & 1 - \sum_{j=1}^{r'} \sum_{i\in\mathbb{I}^{\neq}_{i_j}} \left(\frac{9q_m^{\frac{8}{3}}\delta^{\frac{3}{2}}}{8\cdot 2^{\frac{2}{3}n}} + \frac{31q_m^4\delta + 31q_m^3\delta^{\frac{1}{4}}}{2\cdot 2^{2n}} + \frac{16q_m^4}{2^{3n}}\right) \\
& \hfill (\because \neg\mathsf{Bad3} \wedge \neg\mathsf{Bad5}) \\
\geq & 1 - \left(\frac{9uq_m^{\frac{8}{3}}\delta^{\frac{3}{2}}}{8\cdot 2^{\frac{2}{3}n}} + \frac{47uq_m^3\delta^{\frac{1}{4}}}{2^{2n}}\right) \qquad (43)
\end{aligned}
$$

Conclude the Proof. From Equations (42) and (43), define

$$\epsilon_1 \stackrel{\text{def}}{=} \frac{9uq_m^{\frac{8}{3}}\delta^{\frac{3}{2}}}{8\cdot 2^{\frac{2}{3}n}} + \frac{47uq_m^3\delta^{\frac{1}{4}}}{2^{2n}}$$

and

$$\epsilon_2 \stackrel{\text{def}}{=} \frac{2u^2}{2^{2k}} + \frac{2upq_m\delta}{2^k} + \frac{2u^2q_m^2\delta}{2^k} + 8uq_m^2\delta^{\frac{3}{2}} + \frac{3upq_m}{2^{\frac{n}{2}+k}}.$$

Then by Lemma 1, we conclude that

$$\mathsf{Adv}^{\mathsf{mu\text{-}prf}^*}_{\mathsf{DbHtS}}(u, q_m, p) \leq \frac{2upq_m\delta}{2^k} + \frac{2u^2q_m^2\delta}{2^k} + 10uq_m^2\delta^{\frac{3}{2}} + \frac{3upq_m}{2^{\frac{n}{2}+k}} + \frac{2u^2}{2^{2k}} + \frac{47uq_m^3\delta^{\frac{1}{4}}}{2^{2n}}.$$

References

[1] Bellare, M., Impagliazzo, R.: A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. Cryptology ePrint Archive, Report 1999/024 (1999), https://eprint.iacr.org/1999/024 3

[2] Bellare, M., Desai, A., Jokipii, E., Rogaway, P.: A concrete security treatment of symmetric encryption. In: 38th FOCS. pp. 394–403 3
[3] Bellare, M., Guérin, R., Rogaway, P.: XOR MACs: New methods for message authentication using finite pseudorandom functions. In: CRYPTO'95. LNCS, vol. 963, pp. 15–28 3
[4] Bellare, M., Kilian, J., Rogaway, P.: The security of cipher block chaining. In: CRYPTO'94. LNCS, vol. 839, pp. 341–358 3
[5] Bellare, M., Krovetz, T., Rogaway, P.: Luby-Rackoff backwards: Increasing security by making block ciphers non-invertible. In: EUROCRYPT'98. LNCS, vol. 1403, pp. 266–280 3
[6] Bellare, M., Rogaway, P.: The security of triple encryption and a framework for code-based game-playing proofs. In: EUROCRYPT 2006. LNCS, vol. 4004, pp. 409–426 3
[7] Bernstein, D.J.: How to stretch random functions: The security of protected counter sums. Journal of Cryptology **12**(3), 185–192 3
[8] Bhattacharya, S., Nandi, M.: Full indifferentiable security of the xor of two or more random permutations using the χ^2 method. In: EUROCRYPT 2018, Part I. LNCS, vol. 10820, pp. 387–412 3
[9] Chang, D., Nandi, M.: A short proof of the PRP/PRF switching lemma. Cryptology ePrint Archive, Report 2008/078 (2008), https://eprint.iacr.org/2008/078 3
[10] Chen, S., Steinberger, J.P.: Tight security bounds for key-alternating ciphers. In: EUROCRYPT 2014. LNCS, vol. 8441, pp. 327–350 11
[11] Chen, Y.L., Choi, W., Lee, C.: Improved multi-user security using the squared-ratio method. In: CRYPTO 2023, to apear 2, 3, 4, 5, 6, 8, 9, 11, 42, 50, 51, 64, 66, 67
[12] Chen, Y.L., Mennink, B., Preneel, B.: Categorization of faulty nonce misuse resistant message authentication. In: ASIACRYPT 2021, Part III. LNCS, vol. 13092, pp. 520–550 4, 8
[13] Choi, W., Kim, H., Lee, J., Lee, Y.: Multi-user security of the sum of truncated random permutations. In: ASIACRYPT 2022, Part II. LNCS, vol. 13792, pp. 682–710 3, 6, 42
[14] Choi, W., Lee, B., Lee, J.: Indifferentiability of truncated random permutations. In: ASIACRYPT 2019, Part I. LNCS, vol. 11921, pp. 175–195 3
[15] Choi, W., Lee, B., Lee, J., Lee, Y.: Toward a fully secure authenticated encryption scheme from a pseudorandom permutation. In: ASIACRYPT 2021, Part III. LNCS, vol. 13092, pp. 407–434 3
[16] Choi, W., Lee, B., Lee, Y., Lee, J.: Improved security analysis for nonce-based enhanced hash-then-mask MACs. In: ASIACRYPT 2020, Part I. LNCS, vol. 12491, pp. 697–723 4, 5, 6, 8, 9, 19, 21, 50
[17] Choi, W., Lee, J., Lee, Y.: Building prfs from tprps: Beyond the block and the tweak length bounds. IACR Cryptol. ePrint Arch. p. 918 28
[18] Cogliati, B., Lampe, R., Patarin, J.: The indistinguishability of the XOR of k permutations. In: FSE 2014. LNCS, vol. 8540, pp. 285–302 3
[19] Dai, W., Hoang, V.T., Tessaro, S.: Information-theoretic indistinguishability via the chi-squared method. In: CRYPTO 2017, Part III. LNCS, vol. 10403, pp. 497–523 3, 5, 6, 10
[20] Datta, N., Dutta, A., Nandi, M., Paul, G.: Double-block hash-then-sum: A paradigm for constructing BBB secure PRF. IACR Trans. Symm. Cryptol. **2018**(3), 36–92 4, 5, 68

[21] Datta, N., Dutta, A., Nandi, M., Talnikar, S.: Tight multi-user security bound of dbhts. IACR Trans. Symmetric Cryptol. **2023**(1), 192–223, https://doi.org/10.46586/tosc.v2023.i1.192-223 5, 6, 7, 9, 68, 75, 80

[22] Dutta, A., Nandi, M., Saha, A.: Proof of mirror theory for $\xi_{\max} = 2$. IEEE Trans. Inf. Theory **68**(9), 6218–6232 3

[23] Dutta, A., Nandi, M., Talnikar, S.: Beyond birthday bound secure MAC in faulty nonce model. In: EUROCRYPT 2019, Part I. LNCS, vol. 11476, pp. 437–466 4, 5, 49

[24] Gilboa, S., Gueron, S., Morris, B.: How many queries are needed to distinguish a truncated random permutation from a random function? Journal of Cryptology **31**(1), 162–171 3

[25] Gunsing, A., Mennink, B.: The summation-truncation hybrid: Reusing discarded bits for free. In: CRYPTO 2020, Part I. LNCS, vol. 12170, pp. 187–217 3

[26] Hall, C., Wagner, D., Kelsey, J., Schneier, B.: Building PRFs from PRPs. In: CRYPTO'98. LNCS, vol. 1462, pp. 370–389 3

[27] Impagliazzo, R., Rudich, S.: Limits on the provable consequences of one-way permutations. In: CRYPTO'88. LNCS, vol. 403, pp. 8–26 3

[28] Jha, A., Nandi, M.: Tight security of cascaded LRW2. J. Cryptol. **33**(3), 1272–1317, https://doi.org/10.1007/s00145-020-09347-y 7

[29] Kim, S., Lee, B., Lee, J.: Tight security bounds for double-block hash-then-sum MACs. In: EUROCRYPT 2020, Part I. LNCS, vol. 12105, pp. 435–465 5

[30] Landecker, W., Shrimpton, T., Terashima, R.S.: Tweakable blockciphers with beyond birthday-bound security. In: CRYPTO 2012. LNCS, vol. 7417, pp. 14–30 7

[31] Lee, J.: Indifferentiability of the sum of random permutations toward optimal security. IEEE Trans. Inf. Theory **63**(6), 4050–4054, https://doi.org/10.1109/TIT.2017.2679757 3

[32] Lucks, S.: The sum of PRPs is a secure PRF. In: EUROCRYPT 2000. LNCS, vol. 1807, pp. 470–484 3

[33] Mennink, B.: Towards tight security of cascaded LRW2. In: Theory of Cryptography - 16th International Conference, TCC 2018, Panaji, India, November 11-14, 2018, Proceedings, Part II. Lecture Notes in Computer Science, vol. 11240, pp. 192–222. https://doi.org/10.1007/978-3-030-03810-6_8 7

[34] Patarin, J.: A proof of security in o(2n) for the xor of two random permutations. In: ICITS 2008. LNCS, vol. 5155, pp. 232–248 3

[35] Patarin, J.: A proof of security in $O(2^n)$ for the xor of two random permutations — proof with the "H_σ technique" —. Cryptology ePrint Archive, Report 2008/010 (2008), https://eprint.iacr.org/2008/010 3

[36] Patarin, J.: Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. Cryptology ePrint Archive, Report 2010/287 (2010), https://eprint.iacr.org/2010/287 3

[37] Shen, Y., Wang, L., Gu, D., Weng, J.: Revisiting the security of DbHtS MACs: Beyond-birthday-bound in the multi-user setting. In: CRYPTO 2021, Part III. LNCS, vol. 12827, pp. 309–336 5, 6, 7, 68, 69, 70, 71, 72, 73