On Overidealizing Ideal Worlds: Xor of Two Permutations and its Applications

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 can lead to underestimating the actual security of certain primitives. As a demonstrating example, $\mathsf{XoP2}$, which relies on two independent random permutations, has been proven to exhibit 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, and demonstrate how the transition to the fine-tuned setting can also yield better privacy/authentication tradeoff in authenticated encryption via the generic composition of encryption and MAC. Our analyses demonstrate that the security of XoP-based and $\mathsf{XoP2}$-based constructions are, in fact, far more similar than what was previously proven.

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. [CRYPTO'23].

Keywords: symmetric-key cryptography, provable security, multi-user security, pseudorandom function, xor of permutations, message authentication code

Table of Contents

1 Introduction

Idealization is a common methodology in cryptography in order to enable provable security treatment of complex constructions. The high-level idea is the following: In order to prove a cryptographic primitive secure we compare it to an *idealization*, often referred to as an *ideal (abstraction) model*, which encapsulates the behavior we would hope the primitive to exhibit. The security of the primitive is then argued by means of the distinguishing ability of a distinguisher—-often referred to as the *adversary*—in telling apart the primitive from its ideal abstraction. Looking ahead, in a typical symmetric-key cryptography scenario—such as the ones considered in this work—where concrete security is of essence, this ability is captured by means of the number of queries to the primitive or idealization that an adversary needs to make in order to achieve certain distinguishing advantage.

The most common (and successful) instance of such idealization in symmetric-key cryptography is block ciphers, which are typically abstracted/idealized as pseudorandom permutations (PRPs). In fact, block ciphers offer an excellent segway to discuss the benefits of idealization for security proofs. In particular, in order to allow simpler and black box security proofs of higher lever constructions making use of block ciphers, one typically abstracts keyed block ciphers as (truly) random permutations (RPs). RPs are an easier-to-use idealization of PRPs—security of PRPs is proven by considering how distinguishable they are from RPs.

The above two-level abstraction allows for a simple and modular analysis of primitives, at times referred to as *generic composition* [38, 7, 8] or *game-hopping* [9]: If primitive P can be idealized as primitive P' and primitive P' can be idealized as primitive P'', then P can be idealized as primitive P''. [3] In fact, it is not hard to show that in such a composition, if P' and P'' can be distinguished with advantage δ_1 by an adversary making q_1 queries, and P and P' can be distinguished with advantage δ_2 by an adversary making q_2 queries, then P and P'' can be distinguished with advantage $\delta_1 + \delta_2$ by an adversary making $q = \min\{q_1, q_2\}$ queries. This offers a generic way to bound the distinguishability of P and P'', and therefore the security of P as an instantiation of P''. We note in passing that this bound is typically loose in terms of both queries and distinguishing advantages. Indeed, in many reductions of primitives to their idealization, one is able to prove much tighter bounds—this will also be the case in reductions considered here. Nonetheless, the above generic bound offers us a good way to introduce our motivating question below.

The above discussion illustrates that choosing the "right" idealization can be beneficial in security proofs that make use of the corresponding primitive. Indeed, over-idealizing P as P', i.e., choosing a P' which is closer to (therefore harder to distinguish from) the ideal P'' might improve δ_2 and q_2 at a cost,

[3] We note that often, idealization compares cryptographic games involving the primitive and its abstraction. However, for simplicity, we now use the language of idealizing primitives, which we find simpler and more intuitive to describe our ideas.

however, on δ_1 and q_1 which makes the reduction of P'' to P looser than what one would get by a less idealized P'.

This motivates the core question of this work, namely,

> Can we improve the tightness of security reductions in symmetric-key cryptography by considering better, more *fine-tuned* ideal worlds for the underlying cryptographic primitive?

We answer this question in the affirmative for a broad class of symmetric-key reductions for primitives that rely on the Xor of two permutations—such as the well-known XoP construction (see below)—including multi-user beyond-birthday-bound (BBB) PRFs, Message Authentication Codes (MACs), and Authenticated Encryptions with Associated Data (AEADs). We note that the work of Dai, Hoang, and Tessaro [22] already serves as a demonstration that the use of more fine-tuned abstractions can vastly improve the security proof of XoP. Albeit, as discussed below, unlike our treatment [22] used such fine-tuning inside the security proof, which, as we discuss below, prevented them from performing a tighter analysis of applications like the one we do in this work.

1.1 Related Work

In the following, we discuss (some of the vast) literature on provable security of symmetric-key cryptographic constructions. This will allow us to better demonstrate how our contributions advance the state of the art.

1.1.1 Block ciphers as the basis for PRFs. Pseudorandom Functions are the cornerstone of many cryptographic constructions. Unfortunately, designing PRFs from scratch, without relying on other cryptographic primitives or number-theoretic assumptions, is a challenging task. On the other hand, as discussed above, a common idealization of block ciphers is to treat them as pseudorandom permutations (PRPs). In fact, the plethora of heavily studied and ingeniously engineered block ciphers, such as AES (which is even implemented in commodity hardware), makes PRPs a convenient ingredient in the design of other cryptographic primitives. And indeed, a PRP is a PRF, which is also a permutation, so one would expect the reduction to be trivial. However, it is well known that using PRPs as PRFs in such a vanilla manner has major drawbacks as it makes the resulting primitives susceptible to birthday attacks [5, 6, 9, 12, 30, 31]; such attacks take advantage of the fact that a PRF is a PRP only up to the birthday bound and it can be distinguished by detecting an output collision. This led to the famous foundational question—namely, the Luby-Rackoff backward problem [6]—how to construct secure beyond-birthday-bound (BBB) PRFs from secure PRPs. Remarkably, even more than two decades after the question first surfaced, with several works addressing it [3, 4, 10, 18], there remains broad scope for exploration.

The Xor of Two Permutations as a BBB PRF. Bellare, Krovetz, and Rogaway [6] and Hall et al. [30] pioneered the investigation of constructing beyond-birthday-bound secure PRFs from PRPs, which has since attracted considerable attention [6, 30, 40, 41, 35, 22, 11, 27, 17, 28, 14]. 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

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

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

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

After the initial introduction of XoP construction [6, 30], several studies have built upon and enhanced this groundbreaking work [2, 21, 36, 39]. The most notable advancements include proofs by Dai, Hoang, and Tessaro [22] and Dutta, Nandi, and Saha [25], 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.

Interestingly, the above works demonstrate a gap in the security of the above two primitives: 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. In fact, as discussed below, this becomes ever more prominent in the following two cases: (1) the multi-user setting; and (2) when we want enough security margin. We discuss each of these cases below:

(1) In the multi-user setting, one assumes u instances of the above constructions—each with its own independent key, where the adversary is allowed to make $q = u \cdot q_m$ queries such that q_m queries to each of the above instances. The corresponding ideal primitive is one that offers the adversary parallel query access to u random functions. Choi et al. [16] and Chen, Choi, and Lee [14] improved the multi-user security bound of $\mathsf{XoP2}$. Their result implies that even if there are $O(2^n)$ number of $\mathsf{XoP2}$ instances, i.e., $O(2^n)$ users, $\mathsf{XoP2}$ still enjoys beyond-birthday-bound security. On the other hand, only one query per instance suffices to break PRF security of XoP in the same setting by checking if there is output 0.

(2) Similarly, even for the single-user setting with $n = 128$, if one wants to limit an adversary's advantage to be less than $1/2^{64}$, one can make at most 2^{64} queries to XoP while $\mathsf{XoP2}$ allows 2^{96} queries for the same security level.

1.1.2 Message Authentication Codes.

Message authentication codes (MACs) are the symmetric-key crypto solution to secure authentication. Their goal is to allow two users A and B sharing a secret key to exchange messages in an authenticated manner. As such, the typical security of MACs is captured by means of an ideal game, capturing existential unforgeability, in which the adversary can obtain several MAC tags on messages of his choice and is required to produce a forgery on a message not queried before.

Similarly to PRFs the wide availability of block ciphers (as PRPs) has fuelled a vibrant research investigating constructions of secure MACs using PRPs as building blocks. In fact, there is a very interesting connection between MACs and PRFs. Indeed, any PRF can be directly used as a MAC, by taking the output of the PRF keyed with the secret key of the MAC and evaluated on a message m as the authentication tag for m. Even more intriguingly, it turns out that several of the PRP-based MAC constructions actually produce pseudorandom tags, which means that these constructions can also be seen as alternative PRFs with security analogous to the MAC. In fact, these constructions are very convenient because they can achieve beyond-birthday-bound (BBB) security.

The above state of affairs has led to an extensive body of literature that evaluates security of PRP-based MACs by idealizing them as random functions. This paradigm is often referred to as *PRF-security of MACs* since any indistinguishability statement that proves the MAC produces tags indistinguishable from random can be directly interpreted as a statement about the quality of these MACs when used as a PRF. Notwithstanding, looking ahead in our contributions, abstracting certain PRP-based MACs as random functions is, in fact, an instance of overidealization, which leads to looser bounds on the security of the MAC. Thus, such constructions will greatly benefit from our fine-tuning framework.

Constructing BBB Secure MACs from PRPs. Broadly speaking, there are two main paradigms, called *Nonce-Enhanced Hash-Then-MAC* [26] and *Double-block Hash-then-Sum* [23], to construct BBB secure MACs based on block ciphers. Both paradigms maintain two n-bit internal states and finalize an output by XORing two block cipher evaluations for each state. Intuitively, they can be viewed as a natural extension of XoP construction and share many commonalities with it. We next review the state of the art in proving security for each of these paradigms:

1. Nonce-Enhanced Hash-Then-MAC. Dutta, Nandi, and Talnikar [26] proposed an efficient construction 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:

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

for a n-bit 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. [19] 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

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

using two permutations, which we refer to $\mathsf{nEHtM2}$. This was first considered by Chen, Mennink, and Preneel [15], showing the single-user PRF security of this variant up to $O(2^{3n/4})$ queries. Chen, Choi, and Lee [14] 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 because of similar attacks to XoP. 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. [23]. Notably, a two-keyed construction of DbHtS based on XoP uses one hash key and one block cipher key. This version of DbHtS [33, 42, 24] is of the form

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

where $\mathsf{H} = (\mathsf{H}^1,\mathsf{H}^2)$ consists of two n-bit hash functions $\mathsf{H}^1,\mathsf{H}^2 : \mathcal{K}_h \times \mathcal{M} \rightarrow \{0,1\}^n$ and $\mathsf{E} : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$ is a block cipher.

In the recent improved security analysis of DbHtS [33, 42, 24], Shen et al. [42] proved a multi-user security bound $O\left(\frac{\ell q^3}{2^{2n}} + \frac{qp^2}{2^{2k}}\right)$ in the ideal cipher model for the two-keyed DbHtS, when ℓ is the maximum length of messages, a $\frac{\ell}{2^n}$-universal (and regular) hash is used, and an adversary makes at most q construction queries and p primitive queries. The above security bound assumes that the hash function $\mathsf{H} = (\mathsf{H}^1,\mathsf{H}^2)$ is constructed from the block cipher E and is cross-collision resistant [4], namely for every $M, M' \in \mathcal{M}$,

$$\Pr_{K_{h,1},K_{h,2} \xleftarrow{\$} \mathcal{K}}\left[\mathsf{H}^1_{K_{h,1}}(M) = \mathsf{H}^2_{K_{h,2}}(M')\right] = 0.$$

Later, two-keyed DbHtS are proven to be secure with a better bound $O\left(\frac{\ell q^{4/3}}{2^n} + \frac{pq^{1/3}}{2^k}\right)$ by Datta et al. [24] in terms of threshold number of queries. Their analysis assumes the hash function is cross-collision resistant but not constructed from the block cipher.

[4] The latest version of [42] introduces two variables ϵ_3, ϵ_4 in their security bound to fix the flaw pointed out by [24]. The value of ϵ_3, ϵ_4 depends on concrete hash function constructions and sometimes can be birthday-bound [29]. To make an easier comparison, we simplified their security bound by assuming H is cross-collision resistant, which leads to $\epsilon_3 = \epsilon_4 = 0$.

1.1.3 Generic Composition and Authenticated Encryption.

Namprempre et al. [38] put forth a framework for combining secure symmetric-key encryption with PRF-secure MACs for constructing and analyzing AEAD schemes. The relevant indistinguishability bounds are derived from the security of the encryption scheme and the PRF security of the MAC. As we shall show, this instance of generic composition of encryption and authentication can serve as an example to demonstrate how fine-tuning the idealization of MACs to tighter abstract the properties of PRP-based MACs, like the ones discussed above, allows for a fine-grained treatment of the trade-offs between authentication and privacy, and has the potential to yield overall tighter analyses of AEADs. We refer to Section 3.2 for details.

1.2 Our Contributions

In the following, we discuss our core contribution and put them in context with the related literature.

As discussed above, the starting point for out work is the observation that in PRP-based constructions as above, there is an (as we argue) unnecessary loss in the security analyses that can be gained back by fine-tuning the corresponding ideal worlds. Intuitively, the source for this loss can be demonstrated by looking at the XoP construction: Since the inputs to the two PRPs are domain-separated, the resulting PRF never outputs the all-zero string **0**. As such, a PRF with **0** as part of its output domain is an overidealization of XoP. Using such an overidealization makes it necessary to account in the analysis for the event that the ideal world outputs **0**, which is an event that never occurs in the XoP construction.

By fine-tuning the ideal world (for XoP) to exclude the output **0**, we get a primitive closer to what XoP instantiates and can therefore give tighter reductions, which, among others, demonstrate that the gap in the multi-user setting discussion in Section 5 is mostly due to the above overidealization. Looking ahead, it turns out that such fine-tuning improves the analysis of several results that rely on BBB PRFs, including the ones discussed above.

For instance, 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, Choi, and Lee [14], resulted to the same security bound of XoP2 proven there.

Interestingly, as will become apparent in our analysis, our ideal-world fine-tuning can yield improvements even over the proof of [22], which already considers an intermediate world where **0** is removed from the output of the PRF. Indeed, unlike in our treatment, the security proof from [22] could not surpass the bounds $O\left(\frac{q}{2^n}\right)$ due to the presence of a trivial bad event that outputs **0** in the vanilla ideal world for PRF. Avoiding this obstacle via our fine-tuning leads, as we show, to immediate improvement of security bounds of XoP-based constructions such as nEHtM [26, 19, 14] by the rid of the overestimated assumption. From this observation, we also newly introduce a variant of DbHtS [23, 33, 42, 24], which uses one block cipher key and domain separation.

Similarly to the above-improved analysis for multi-user XoP, we can get improvement on 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. Furthermore, we demonstrate how fine-tuning the ideal world for such MACs can serve as an avenue for better privacy/authentication tradeoffs (and potentially tighter security analysis) of authenticated encryption using the generic composition framework of Namprempre et al. [38].

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.

In the following, we give details for our results on fine-tuned (less idealized) ideal worlds for XoP, nEHtM, and (a variant of) DbHtS and their applications. 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 the adversary.

1.2.1 Revisiting the security of (multi-user) XoP. We show that the "fine-tuned" multi-user PRF security bound of XoP from the random ideal world without outputting zero can be $O\left(\frac{u^{0.5}{q_m}^2}{2^{2n}}\right)$ via the Squared-ratio method [14] where the same security bound for XoP2 was also proven in [14]. 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 $\mathbf{0}$.

1.2.2 Revisiting the security of (multi-user) 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 [14] while using less key size with graceful security degradation under nonce misuse, resolving the open question posed in [14]. 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 $O\left(\frac{uq_m^2}{2^{1.5n}}\right)$ [19]. 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 [14] is buggy (see Appendices D.3 and D.4), 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 [14] 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 [14] 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.

1.2.3 Revisiting the security of (multi-user) DbHtS. Subsequently, we explore the multi-user MAC security of DbHtS. Our main targets are the variants of [24, 42] using the domain separation, whose formal definition can be found in Section 7. 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 [42], 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 [24], 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. We refer readers to Appendix E.1 for additional pictorial comparisons in different settings.

Remark 1 (Strong hash functions). The strong properties of hash functions used in the tighter security proofs are likely satisfied when we use block-cipher-based hash functions. On the other hand, the polynomial hash functions do not satisfy them, highlighting the specific point in the current proofs. We leave an open

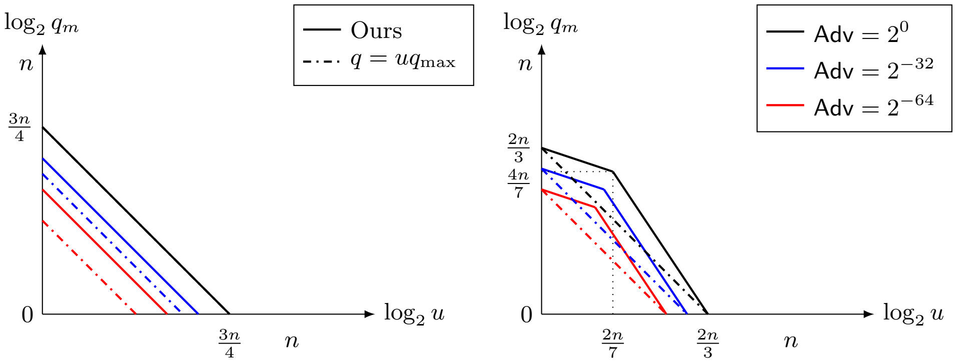

Fig. 1: Comparison of the DbHtS's security bounds (in terms of the threshold number of queries per user) as functions of $\log_2 u$. The black line represents the bound where adversary advantage $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{DbHtS}}(u, q_m, p) = 2^0$; the blue line is for $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{DbHtS}}(u, q_m, p) = 2^{-32}$; and the red line is for $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{DbHtS}}(u, q_m, p) = 2^{-64}$. 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 9 with Theorem 1 from [24]. Two bounds overlap. The right figure compares our Theorem 8 with Theorem 1 from [42], where we set $\epsilon_3, \epsilon_4 = 0$. We set $p = q, k = n, \delta = 2^{-n}$, and neglect l and the logarithmic term of n in all graphs.

question of whether polyhash can be used to obtain similar tight security or admit the matching attacks, especially due to the lack of such properties.

This situation is reminiscent of the recent advances in the cascaded $\mathsf{LRW2}$ [34], or $\mathsf{CLRW2}$. Mennink [37] presented an attack on $\mathsf{CLRW2}$ and showed that matching security bound under several assumptions, including stronger hash functions very similar to ours. Subsequently, Jha and Nandi [32] eliminated those assumptions and developed a new tool for the hash functions. In turn, these tools are frequently used in the later works [14, 19, 15] as well as this work.

1.2.4 Revisiting the security of generic composition (for AEADs)

Last but not least, we demonstrate that our proposal of replacing the idealization of MAC as PRF with the idealization of PRF without the **0** in the codomain—we will refer to this as *PRF** and the resulting security notions as *PRF*-security*—not only yields improvements in the analysis of the MAC security but also leads to a potential avenue to improve the analysis of generic composition of encryption and MAC towards authenticating encryption. In particular, we show that by replacing the PRF with PRF* in the analysis of Namprempre et al. [38] (and plugging in out tighter analysis of PRF*-security for PRP-based MACs), we can make δ_1 (the advantage of the adversary in attacking the MAC) small while keeping the change on δ_2 (the adversary's advantage in attacking the encryption) very low, at least for the CTR mode. Hence, the transition from PRF security of MACs to PRF* security of MACs allows us to improve security with respect to authentication at a minimal loss in security with respect to privacy. The caveat here is that in the CTR mode $\delta_2 \gg \delta_1$ hence δ_2 is, in fact, the dominating factor that leverages the overall gains of transitioning to a PRF*-security analysis. Such gains would appear if we used an encryption mode for which $\delta_1 > \delta_2$ (or at least $\delta_1 \approx \delta_2$). Unfortunately, there are only a few encryption mode candidates with this property, and we were unable to provide rigorous security proof (as for the CTR mode) using PRF* for these modes—the analysis becomes overly complex. We conjecture that such AEADs which we find a very interesting future research direction.

1.3 Organization of the Paper

The remainder of the paper is organized as follows. In Section 2, we introduce some basic notation and preliminaries. In Section 3, we present the formal definition of the fine-tuned pseudorandom function, denoted by PRF*, multi-user PRF* security, and MAC security. We then discuss generic compositions from PRF*. In Section 4, we develop and state the Mirror Theory tailored for the fine-tuned ideal world. The Mirror Theory will later be used in the proof of multi-user PRF* security of XoP and multi-user MAC security of the nEHtM and DbHtS. Then, we show how using the fine-tuned ideal world could improve the security analysis in multiple applications. Specifically, Section 5 states and proves the improved multi-user PRF* security of XoP. Section 6 and Section 7 state and prove the improved multi-user MAC security for the nEHtM and DbHtS, separately.

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 $\mathcal{X} \sqcup \mathcal{Y}$. For any positive integer i, and $a_1, \ldots, a_i, b \in \{0,1\}^n$, We denote $\{a_1, \ldots, a_i\} \oplus b \overset{\text{def}}{=} \{a_1 \oplus b, \ldots, a_i \oplus b\}$

In the proofs in our paper, we denote $\mathcal{S}_1^i$ and $\mathcal{S}_0^i$ as a random system of the real world and the ideal world for an i-th user. The random variable $z \in \Omega$ sampled from $\mathcal{S}_1^i$ or $\mathcal{S}_0^i$ is called a transcript, which consists of query-answer pairs between an adversary and a given random system based on a distinguishing game described later. z is *attainable* if the probability of obtaining z from $\mathcal{S}_0^i$ is non-zero. 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.

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 Squared-Ratio Method

This method was first introduced in [14]. For multi-user security, we assume a sequence of random systems $(\mathcal{S}^1, \ldots, \mathcal{S}^u)$ 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. 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 a system $\mathcal{S}$. Let

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

and let

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

for $\mathbf{z} \in \Omega^i$. We omit i when $i = q$.

Theorem 1 ([14]). *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 $Z_{\mathcal{S}_0^1}$.*

3 Fine-tuned Pseudorandom Functions and Applications

This section introduces the notion of a family of functions, which we denote by PRF*, that is indistinguishable from the truly random functions sampled from $\mathsf{Func}^*(m, n)$, a set of functions from $\{0,1\}^m$ to $\{0,1\}^n \setminus \{0^n\}$. We note that PRF* are indeed pseudorandom functions. Still, the codomain is chosen slightly differently from the usual pseudorandom functions in symmetric-key cryptography, which we call original or *full-domain* (following the full-domain hash functions). We also note that this notion is considered in [22] in the middle of security proofs under the name of "normalized" ideal worlds.

The choice of codomain $\{0,1\}^n \setminus \{0^n\}$ for PRF* makes a potential doubt about their practical usability. Despite this concern, we prove this is not true for many settings. The latter part of this section argues this point rigorously: PRF* as is (or with some properties that the original proof also assumed) can be used for many applications instead of the full-domain PRF, albeit with the codomain without 0^n. This proves the tighter security of many applications of the XoP paradigm, combined with the results of this paper where some XoP constructions are shown to have stronger security as PRF* than that as the full-domain PRF.

3.1 Fine-tuned Pseudorandom Functions and MACs

We present the formal definition of multi-user PRF* and nonce-based multi-user MAC. We do *not* change the security notion for the MAC, but a conventional construction using PRF works well for PRF*.

Fine-tuned Pseudorandom Functions. We consider the multi-user PRF* security throughout this paper for the number of users u. The single-user security coincides with the setting of $u = 1$.

Definition 1. *Let* $\mathsf{C} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^m$ *be a keyed function with the key space* $\mathcal{K}$*. Let* u *be the number of users. For a distinguisher* $\mathcal{A}$*, we define the PRF* advantage of* $\mathcal{A}$*, denoted by* $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(\mathcal{A})$*, is defined by*

$$\left| \Pr_{K_1,\ldots,K_u \leftarrow \mathcal{K}} \left[\mathcal{A}^{\mathsf{C}_{K_1},\ldots,\mathsf{C}_{K_u}} \to 1\right] - \Pr_{\mathsf{F}_1,\ldots,\mathsf{F}_u \leftarrow \mathsf{Func}^*(n,m)} \left[\mathcal{A}^{\mathsf{F}_1,\ldots,\mathsf{F}_u} \to 1\right] \right|.$$

We say that C *is an (information-theoretic)* (ϵ, u, q_m)*-PRF* if* $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(\mathcal{A}) \leq \epsilon$ *for all distinguishers for* u *users and making at most* q_m *queries to each oracle. The maximum* ϵ *in the same setting is denoted by* $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(u, q_m)$*.*

We sometimes consider the setting where the oracle used in the construction can be accessed as a global primitive- usually the ideal cipher. We consider the following definition in this setting.

Definition 2. *Let* E *be a (publicly accessible and possibly random) function. Let* $\mathsf{C} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^m$ *be a keyed function, which may depend on* E*, with the key space* $\mathcal{K}$*. Let* u *be the number of users. For a distinguisher* $\mathcal{A}$*, we define the PRF* advantage of* $\mathcal{A}$*, denoted by* $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(\mathcal{A})$*, is defined by*

$$\left| \Pr_{K_1,\ldots,K_u \leftarrow \mathcal{K}, \mathsf{E}} \left[\mathcal{A}^{\mathsf{E},\mathsf{C}_{K_1},\ldots,\mathsf{C}_{K_u}} \to 1\right] - \Pr_{\mathsf{F}_1,\ldots,\mathsf{F}_u \leftarrow \mathsf{Func}^*(n,m), \mathsf{E}} \left[\mathcal{A}^{\mathsf{E},\mathsf{F}_1,\ldots,\mathsf{F}_u} \to 1\right] \right|.$$

We say that C *is an* (ϵ, u, q_m, p)*-PRF* if* $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(\mathcal{A}) \leq \epsilon$ *for all distinguishers* $\mathcal{A}$ *for* u *users, making at most* q_m *queries to each oracle, and making at most* p *queries to* E*. The maximum* ϵ *in the same setting is denoted by* $\mathsf{Adv}_{\mathsf{C}}^{\mathsf{mu\text{-}prf}^*}(u, q_m, p)$*.*

Nonce-based MACs. The MAC scheme consists of two algorithms: MAC and verification. The verification algorithm returns $\top$ ("accept") if the input is valid MAC, and otherwise $\bot$ ("reject"). We consider the multi-user setting, where the $u = 1$ case coincides with the single-user case.

Definition 3. *Let* $\mathcal{K}$*,* $\mathcal{N}$*,* $\mathcal{M}$*, and* $\mathcal{T}$ *be non-empty sets. A nonce-based MAC scheme* MAC *consists of the MAC algorithm* $\mathsf{S} : \mathcal{K} \times \mathcal{N} \times \mathcal{M} \to \mathcal{T}$ *and the verification algorithm* $\mathsf{V} : \mathcal{K} \times \mathcal{N} \times \mathcal{M} \times \mathcal{T} \to \{\top, \bot\}$*. We define* $\mathsf{S}_K(\cdot,\cdot,\cdot) := \mathsf{S}(K,\cdot,\cdot,\cdot)$ *and similarly for* V_K*. We say that* MAC *is correct if* $\mathsf{V}_K(N, M, \mathsf{S}_K(N, M)) = \top$ *holds for any* $(K, N, M) \in \mathcal{K} \times \mathcal{N} \times \mathcal{M}$*.*

Let $K_1, ..., K_u$ *be randomly chosen keys. We consider an algorithm having oracle access to* $\mathsf{S}_{K_1}, ..., \mathsf{S}_{K_u}$ *and* $\mathsf{V}_{K_1}, ..., \mathsf{V}_{K_u}$*, and without loss of generality assume that it never makes the verification query that it received from the MAC query. We define* the MAC advantage *of* $\mathcal{A}$ *against* MAC*, denoted by* $\mathsf{Adv}_{\mathsf{MAC}}^{\mathsf{mu\text{-}mac}}(\mathcal{A})$*, is defined by*

$$\Pr_{K_1,\ldots,K_u \leftarrow \mathcal{K}}\left[\mathsf{V}_{K_i}(N, M, T) = \top \,\middle|\, \mathcal{A}^{\mathsf{S}_{K_1},\ldots,\mathsf{S}_{K_u},\mathsf{V}_{K_1},\ldots,\mathsf{V}_{K_u}} \to (i, N, M, T)\right].$$

If the above event occurs, we say $\mathcal{A}$ forges*.*

A MAC query (N, M) *made by an adversary is called* faulty *if the adversary has already queried the MAC oracle with the same nonce with a different message. We say that* MAC *is* $(\epsilon, u, \mu_m, q_m, v_m)$*-MAC if* $\mathsf{Adv}_{\mathsf{MAC}}^{\mathsf{mu\text{-}mac}}(\mathcal{A}) \leq \epsilon$ *for all adversaries* $\mathcal{A}$ *for* u *users, making at most* q_m *MAC queries, at most* μ_m *faulty queries, and at most* v_m *verification queries to* S_{K_i} *and* V_{K_i} *for each* $i \in [u]$*. The maximum* ϵ *in the same setting is denoted by* $\mathsf{Adv}_{\mathsf{MAC}}^{\mathsf{mu\text{-}mac}}(u, \mu_m, q_m, v_m)$*.*

We sometimes call both the faulty query and the corresponding previous query with the same nonce by a query with a repeated nonce. When $\mu_m = 0$, we say that $\mathcal{A}$ is nonce-respecting.

In this work, we prove the MAC security of some XoP constructions by comparing it with the PRF*-style ideal world of MAC, adapting the conventional strategy through PRF-style ideal worlds. That is, we set $\mathcal{T} = \{0,1\}^t \setminus \{0^t\}$ for some t and consider, for each $i \in [u]$, the random oracles F_i^* sampled from the set of all functions from $\mathcal{N} \times \mathcal{M} \to \mathcal{T}$ and Rej_i that always returns $\bot$. The following lemma provides the formal statement of this idea. We mainly focus on showing the below indistinguishability for MAC security in the other parts of this paper.

Lemma 1. *Let* MAC *satisfies the syntax of MAC schemes for the MAC algorithm* S *and verification algorithm* V*, and for any* (u, μ_m, q_m, v_m)*-distinguisher* $\mathcal{D}$ *it holds that*

$$\left|\Pr_{K_1,\ldots,K_u}\left[\mathcal{D}^{\mathsf{S}_{K_1},\ldots,\mathsf{S}_{K_u},\mathsf{V}_{K_1},\ldots,\mathsf{V}_{K_u}} \to 1\right] - \Pr_{\mathsf{F}_1^*,\ldots,\mathsf{F}_u^*}\left[\mathcal{D}^{\mathsf{F}_1^*,\ldots,\mathsf{F}_u^*,\mathsf{Rej}_1,\ldots,\mathsf{Rej}_u} \to 1\right]\right| \leq \epsilon.$$

Then it holds that $\mathsf{Adv}^{\mathsf{mu\text{-}mac}}_{\mathsf{MAC}}(u, \mu_m, q_m, v_m - 1) \le \epsilon$.

Proof. Suppose that a $(u, \mu_m, q_m, v_m - 1)$-adversary $\mathcal{A}$ has an advantage ϵ'. Consider a distinguish $\mathcal{D}$ that runs $\mathcal{A}$ to obtain the output (i, N, M, T) and returns 1 if and only if $V_{K_i}(N, M, T) = \top$. $\mathcal{D}$ is a (u, μ_m, q_m, v_m)-distinguisher because it makes one additional verification query for the last step, and $\mathcal{D}$ outputs 1 with probability ϵ' in the left world, but always output 0 in the right world. This means $\epsilon' \le \epsilon$, completing the proof. □

3.2 Generic Compositions with PRF*

Namprempre et al. [38] classified and analyzed generic compositions from specifically defined symmetric-key encryption schemes and MAC schemes. In particular, they assumed a given MAC takes a vector input and is PRF secure. This section describes that the PRF* notion for MACs, instead of PRF, suffices for many generic compositions for constructing nonce-based authenticated encryption (nAE) schemes from Namprempre et al. [38]. We particularly focus on the generic compositions using nonce-based encryptions and MACs (Nn family)

The nAE schemes is a tuple $\mathit{\Pi} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ that are defined over a nonempty set of keys $\mathcal{K}$, associated data $\mathcal{A}$, nonce $\mathcal{N}$, messages $\mathcal{M}$, and ciphertexts $\mathcal{C}$. An encryption $\mathcal{E} : \mathcal{K} \times \mathcal{N} \times \mathcal{A} \times \mathcal{M} \to \mathcal{C}$ and decryption $\mathcal{D} : \mathcal{K} \times \mathcal{N} \times \mathcal{A} \times \mathcal{C} \to \mathcal{M} \cup \{\bot\}$ are deterministic. We write $\mathcal{E}_K(\cdot,\cdot,\cdot)$ to denote $\mathcal{E}(K,\cdot,\cdot,\cdot)$ and similarly define $\mathcal{D}_K$. We say $\mathit{\Pi}$ is correct if $\mathcal{D}_K(N, A, \mathcal{E}_K(N, A, M)) = M$ for all K, N, A, M. The security is defined as follows by comparing $\mathit{\Pi}$ with the idealized world.

Definition 4. *Let $\mathit{\Pi} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be an nAE scheme with the non-empty sets $\mathcal{K}, \mathcal{A}, \mathcal{N}, \mathcal{M}, \mathcal{C}$ as described above. The nAE advantage of $\mathit{\Pi}$ is defined by*

$$\mathsf{Adv}^{\mathsf{nAE}}_{\mathit{\Pi}}(q, d) := \max_{\mathcal{A}} \left| \Pr_{K \leftarrow \mathcal{K}}\left[\mathcal{A}^{\mathcal{E}_K, \mathcal{D}_K} \to 1\right] - \Pr_{\mathsf{F}}\left[\mathcal{A}^{\mathsf{F},\mathsf{Rej}} \to 1\right] \right|$$

where the maximum is over the algorithms $\mathcal{A}$ that makes q and d queries to the left and right oracles, respectively, and never making the query (N, A, C) to the right oracle after obtaining C from a previous query (N, A, M) to the left oracle.

This is almost identical to the definition in [38], except that the sets are not explicitly fixed as $\{0,1\}^t$ for some t.[5] In the same vein as Lemma 1, it suffices to make $\mathit{\Pi}$ IND-CPA and authenticated.[6] We proceed with some explicit generic compositions with PRF* below.

N1 SCHEME. Let $\mathsf{SKE} = (\mathsf{E}, \mathsf{D})$ and $\mathsf{MAC} = (\mathsf{S}, \mathsf{V})$ be nonce-based symmetric key encryption and MAC with key spaces $\mathcal{K}_{\mathsf{SKE}}, \mathcal{K}_{\mathsf{MAC}}$, respectively. The codomain of S is denoted by $\mathcal{T}$. Let $\mathcal{K} = \mathcal{K}_{\mathsf{SKE}} \times \mathcal{K}_{\mathsf{MAC}}$. An N1 scheme $\mathit{\Pi} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ is defined as follows.

[5] We also excluded the invalid inputs for simplicity.

[6] We refer the IND-CCA3 definition in [43] for a related discussion. Our security notion (and one in [38]) suffices for proving the security requirements for the authenticated encryption, but the opposite direction does not hold.

- $\mathcal{E}_{K,L}(N, A, M)$: outputs $\mathsf{E}_K(N, M) \| \mathsf{S}_L(N, A, M)$.
- $\mathcal{D}_{K,L}(N, A, C)$: Parses $C = C_{\mathsf{SKE}} \| T_{\mathsf{MAC}}$ and computes $M' = \mathsf{D}_K(N, C_{\mathsf{SKE}})$. Outputs M' if $T_{\mathsf{MAC}} = \mathsf{S}_L(N, A, M')$, otherwise $\perp$.

The security proof of the N1 scheme with PRF*-style MACs is almost identical to the proof of [38, Lemma 4]. Roughly, we can prove the following inequality:

$$\mathsf{Adv}^{\mathsf{nAE}}_{\Pi}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{prf}^*}_{\mathsf{V}}(\mathcal{B}(\mathcal{A})) + \mathsf{Adv}^{\mathsf{auth}}_{\tilde{\Pi}}(\mathcal{A}) + \mathsf{Adv}^{\mathsf{nE}}_{\mathsf{E}}(\mathcal{D}(\mathcal{A}))$$

where $\mathcal{B}(\mathcal{A})$ and $\mathcal{D}(\mathcal{A})$ be black-box reductions identical to original one, and $\tilde{\Pi}$ be the composition $\Pi[\mathsf{SKE}, \mathsf{RF}^*]$ that uses a random function sampled from $\mathsf{Func}^*(m, n)$ as a MAC scheme. nE means the security of nonce-based encryption defined in the original paper. To prove this, it suffices to modify [38, Equation 5] for PRF*. The other parts are unaffected by this change (while the authenticity game shall be "fine-tuned" accordingly). This proof can be generalized to N2 and N3 schemes. This inequality indeed provides us with a good trade-off between authenticity ($\mathsf{Adv}^{\mathsf{auth}}_{\tilde{\Pi}}(\mathcal{A})$) and privacy ($\mathsf{Adv}^{\mathsf{nE}}_{\mathsf{E}}(\mathcal{D}(\mathcal{A}))$), where the privacy decrease is negligible at least in case of CTR mode. Unfortunately, there are only a few encryption mode candidates BBB secure based on block ciphers [18, 1], and we were unable to provide rigorous security proof (as for the CTR mode below) using PRF* for these modes—the analysis becomes overly complex. However, a rough observation implies that if an IV-based encryption scheme is secure up to its IV collisions, the collision probability difference between IVs generated by PRF and PRF* is negligible. See below for the detail.

IV-based Encryptions and More. In most IV-based encryption (ivE) schemes, such as CTR mode, the security can be reduced to the probability of colliding IVs, which is $\binom{q}{2}/2^n$ when IVs are sampled from $\{0,1\}^n$. These IVs can be generated as the PRF outputs. When we use PRF* instead, the IVs are randomly chosen from $\{0,1\}^n \backslash \{0^n\}$, and the probability becomes $\binom{q}{2}/(2^n - 1)$ where the difference, $O(q^2/2^{2n})$, is negligible to compare with collision probabilities. In such cases, the cardinality of the space (2^n for PRF and $2^n - 1$ for PRF*) is the only matter. This observation can be applied further in various situations, e.g., using PRF* as a key-derivation function.

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 set $\mathcal{P} = \{P_1, ..., P_r\}$ is 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 system of equations $\Gamma^{=} \overset{\text{def}}{=} (P_{\gamma_i} \oplus P_{\gamma'_i} = \lambda_i)_{i \in [q]}$

holds. Similarly, a sequence of constants $(\mu_1, ..., \mu_p) \in (\{0,1\}^n)^p$ and indices $\sigma_1, \dots, \sigma_p, \sigma'_1, \dots, \sigma'_p \in [r]$ determine the system of inequations where $\sigma_i \neq \sigma'_{i'}$ for any $i, i' \in [r]$: $\Gamma^{\neq} \overset{\text{def}}{=} (P_{\sigma_i} \oplus P_{\sigma'_i} = \mu_i)_{i \in [p]}$. When the variables in $\mathcal{P}$ are assigned 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

$$\mathcal{V} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup ... \sqcup \mathcal{C}_{\alpha+\beta} \sqcup \mathcal{D} \tag{1}$$

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 representative, we define $\lambda_i(W) \overset{\text{def}}{=} \lambda(V_i, W)$ 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 $i \neq j$[7] 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

$$V_i \oplus \lambda_i(W) \neq V_j \oplus \lambda_j(W'). \tag{2}$$

Now we consider the edges in $\mathcal{E}^{\neq}$ with respect to Equation (1). 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 $V_i \neq V_j \oplus \nu$. 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 (2). Also, if two non-equations in $\mathcal{E}^{\neq}_{i,j}$ 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 $v_{i,j} = |\mathcal{E}^{\neq}_{i,j}|$ 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 $i \neq j$:

$$\{\lambda_i(W) \oplus \lambda_j(W')\}_{(W,W') \in \mathcal{C}_i \times \mathcal{C}_j} \cup \{\mu \oplus \lambda_i(V) \oplus \lambda_j(V')\}_{(V,V') \in \mathcal{E}^{\neq}_{i,j} : [V \oplus V' \neq \mu] \in \Gamma^{\neq}},$$

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.

[7] 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.

- 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 $i \in [\alpha + \beta]$:

$$\mathcal{R}_i \overset{\text{def}}{=} \left\{(V_1, V_1', V_2, V_2') \in {\mathcal{C}_i}^{*2} \times {\mathcal{C}_j}^{*2} \,\middle|\, j < i \text{ and } \lambda(V_1, V_1') = \lambda(V_2, V_2')\right\}. \quad (3)$$

Theorem 2 (Mirror Theory for $\xi_{max} > 2$). *Let $\mathcal{G}$ be 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

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

we have

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

This theorem combines Theorem 4 and Theorem 10 — the lower and upper bounds of

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

that are proven in Appendix Appendices B.1 and B.3, and the final statement is from $e^x \leq 1 + 2x$ for $x \leq 1$.

Below, we give a Mirror theory for equation systems with all component sizes 2. Theorem 3 is used in a multi-user security proof of XoP discussed in Section 5.

Theorem 3 (Mirror Theory with $\xi_{max} = 2$). *Let $\mathcal{G}$ be 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 $q \leq \frac{N}{13}$ and $q_c = 0$, it holds that*

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} - 1\right| \leq \exp\left(\frac{3\sum_{i=1}^{q} |\mathcal{R}_i|}{N} + \frac{12q^2}{N^2} + \frac{10(n+1)^2}{N}\right) - 1.$$

Further, if $\frac{3\sum_{i=1}^{q}|\mathcal{R}_i|}{N} + \frac{12q^2}{N^2} + \frac{10(n+1)^2}{N} < 1$, *it holds*

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} - 1\right| \leq \frac{6\sum_{i=1}^{q}|\mathcal{R}_i|}{N} + \frac{24q^2}{N^2} + \frac{20(n+1)^2}{N}.$$

Proof. By Theorem 11 (see Appendix B.4 for details), we have

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

and by Theorem 5 (deferred to the end of this section), we have

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

Since $\max\{\frac{128q^3}{N^3} + \frac{2q^2}{N^2}, \frac{147q^3}{N^3}\} \leq \frac{12q^2}{N^2}$, and $\frac{8(n+1)^3}{3N^2} \leq \frac{10(n+1)^2}{N}$, we conclude with

$$\left|\frac{h(\mathcal{G})(N-1)^q}{(N)_{C_{\alpha+\beta}}} - 1\right| \leq \exp\left(\frac{3\sum_{i=1}^{q}|\mathcal{R}_i|}{N} + \frac{12q^2}{N^2} + \frac{10(n+1)^2}{N}\right) - 1.$$

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

Theorems 4 and 5 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 5 is used in the proof of Theorem 8 and Theorem 4 is used in the proof of Theorem 9. Note that both theorems are parts of Theorem 2 and 3. The proofs can be found in each proof section.

Theorem 4 (Lower Bound Mirror Theory for $\xi_{max} > 2$). *Assume that* $8q \leq N$. *It holds that*

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

Theorem 5 (Lower Bound Mirror Theory for $\xi_{max} = 2$). *Let* $q \leq \frac{N}{13}$ *and* $q_c = 0$. *Then, it holds that*

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

5 Multi-User Security of XoP

We show 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 and $x \in \{0,1\}^{n-1}$. The result is obtained by following the paradigm of Chen, Choi, and Lee [14], namely the Squared-ratio method. The authors showed that multi-user PRF security of $\mathsf{XoP2}$ is bounded by $O\left(u^{0.5}{q_m}^2/2^{2n}\right)$, which in turn the bound is same to multi-user PRF* security bound of XoP by the following theorem.

Theorem 6. *Let n, u, and q_m be positive integers such that $n > 12$ and $q_m \leq \frac{2^n}{4n}$. Then, it holds*

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

The proof is deferred to Appendix C. Thus security of XoP and $\mathsf{XoP2}$ constructions are, in fact, far more similar than what was previously proven. We also noted that Dai, Hoang, and Tessaro [22] already consider an intermediate world where $\mathbf{0}$ is removed from the output of the PRF while they introduced the fine-tuned ideal world as an intermediate world when they proved (single-user) security of XoP using the Chi-squared method. It shows the compatibility of the fine-tuned model with the Chi-squared method and that XoP and $\mathsf{XoP2}$ enjoy the same security bound in the fine-tuned setting regardless of proving tools.

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 [26]. 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:

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

Recall Definition 3 for the MAC security. 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. We let u be the number of users, q_m, μ_m, v_m the number of maximum MAC, faulty, and verification queries for each $i \in [u]$. The main result of this section is summarized as follows.

Theorem 7. *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. 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:

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

When $q_m \approx 2^{2n/3}$ and $u \approx 2^{n/3}$, the threshold number of the maximum faulty query is $\mu_m \approx 2^{n/3}$ in this bound. On the other hand, the previous best bound [19] with the hybrid argument only gives the threshold about $\mu_m \approx 2^{n/6}$. 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 Appendix D.2. Figure 2 shows the graphical comparison between our bounds and the previous bounds [14, 19] in this setting.

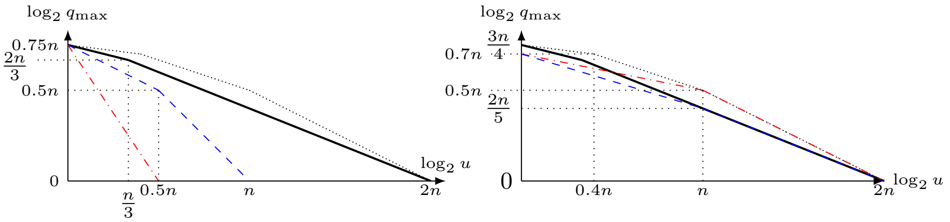

Fig. 2: Comparison of the multi-user security bounds (in terms of the threshold number of queries per user) as functions of $\log_2 u$. We neglect the polynomial terms of ℓ and $\log n$ in the graphs. We assume $v_m = \mu_m = 0$ for a fair comparison. The solid line represents our bounds in both graphs. 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 [14] with our correction in Appendix D.4. The red dash-dotted line in the right graph corresponds to the claimed security bound in [14], 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.

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

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

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:

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

In the remainder of this section, we prove Theorem 7 using Theorems 1 and 2. The proof sketch of the nonce-respecting setting can be found in Appendix D.2. The stronger bound with a pairwise δ-AXU and the discussion on the previous $\mathsf{nEHtM2}$ security proof [14] is placed in Appendix D.4.

Before starting the proof, some observations are in order. First, we always assume that $q_m \leq \frac{2^{3n/4}}{8} \leq \frac{2^n}{256}$, $\mu_m \leq \frac{2^{0.5n}}{12\sqrt{n}}, \frac{2^n}{32q_m}$, $nuq_m\delta < 2^n$, and $v_m \leq \frac{2^n}{128}$, 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: $\tau_m = (N_i, M_i, T_i)_{1 \leq i \leq q_m}$ and $\tau_v = \left(N_j', M_j', T_j', b_j'\right)_{1 \leq j \leq v_m}$ 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 $X_i \stackrel{\text{def}}{=} \mathsf{H}_{K_h}(M_i) \oplus N_i$ and $X_j' \stackrel{\text{def}}{=} \mathsf{H}_{K_h}(M_j') \oplus N_j'$ 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.[8] 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 (4)), 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

$$\mathsf{P}(1\|X_{i_1}) - \mathsf{P}(0\|N_{i_1}) = \mathsf{P}(0\|N_{i_2}) - ..., \text{ or } X_{i_1} - N_{i_1} = N_{i_2} - ... \qquad (4)$$

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 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 $L_1, L_2 \leq \min\left(\frac{2^n}{32q_m}, \frac{2^{0.5n}}{24\sqrt{n}}\right)$.

- 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.

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

  - 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}: $\exists$ an n-collision of length 1 trails.
  - bad_{5b}: $\exists$ an n-collision of length 2 N-trails.
  - bad_{5c}: $\exists$ an n-collision of length 2 X-trails.
  - bad_{5d}: $\exists$ an 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 Appendix D.1.

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 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.

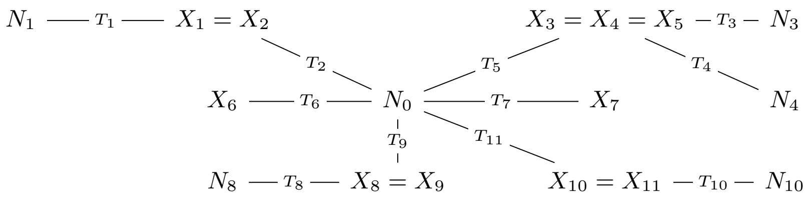

Fig. 3: An example of tree$_{\geq 3}$. In each edge, the tag T_i corresponds to the query $\mathsf{P}(0\|N_i) \oplus \mathsf{P}(1\|X_i)$, where X_i and N_i are written in each vertex. The root is N_0, which is equal to N_2, N_3, N_6, N_7, N_9, and N_{11}. $N_1, N_3, N_4, X_6, X_7, N_8$, and N_{10} are 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.* (1)*. The number of trails in* $\mathcal{C}_1, ..., \mathcal{C}_\alpha$ *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 an* ℓ*-collision* $\{\gamma_1, ..., \gamma_\ell\}$*, there exists a set of indices* $\{i_1, ..., i_\ell\}$ *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:

$$\epsilon_2 := \frac{\ell(7\mu_m^2 + 2v_m)}{2^n} + \frac{3\ell q_m^2 L_1}{2^{2n}} + \frac{3\ell \mu_m q_m}{2^n L_1} + \frac{3\ell q_m^2}{2^n L_2^2} + \frac{372\ell q_m^4}{2^{3n}}. \tag{5}$$

The detailed analysis is deferred to Appendix D.1.

Good transcript analysis. We now assume that the transcript is good, i.e., no bad events occur. Looking ahead, we will use Theorems 1 and 2 to derive the security bound. Recall the notations in Section 4. In particular, there are $\alpha+\beta$ components $\mathcal{C}_i$ with the number of vertices $c_i \geq 2$ for $i \in [\alpha+\beta]$ Equation (1). We divide $\mathcal{R}_i$ into two sets:

$$\mathcal{S}_i \overset{\text{def}}{=} \left\{(V_1, V_1', V_2, V_2') \in \mathcal{R}_i \,\middle|\, \overline{V_1V_1'}, \overline{V_2V_2'} \in \mathcal{E}\right\}, \mathcal{D}_i \overset{\text{def}}{=} \mathcal{R}_i \setminus \mathcal{S}_i.$$

Let $S \overset{\text{def}}{=} \sum_{i=1}^{\alpha+\beta} |\mathcal{S}_i|$. 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 4 to obtain

$$\mathbf{Ex}[S] \leq \frac{q_m^2}{2B}, \quad \mathbf{Ex}\left[S^2\right] \leq \begin{cases} \frac{q_m^2}{B} & \text{if } \frac{q_m^2}{2} < B, \\ \frac{q_m^4}{2B^2} & \text{otherwise}, \end{cases} \tag{6}$$

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

Let $C \overset{\text{def}}{=} \sum_{i=1}^{\alpha} c_i^2$. To count the other terms, we first consider $\mathcal{D}_i$ for $i \leq \alpha$. For each $(V_1, V_1') \in \mathcal{C}_i^{*2}$, $\neg\mathsf{bad}_5$ asserts that there are at most $4n$ different trails corresponding to $(V_2, V_2') \in \mathcal{V}^{*2}$ colliding with the trail for (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 $j \leq \alpha$, because it is included in $\mathcal{S}_i$ otherwise. For each $(V_2, V_2') \in \mathcal{C}_j^{*2}$ for $j \leq \alpha$, 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_1'. Therefore, we have

$$\sum_{i=1}^{\alpha+\beta} |\mathcal{D}_i| \leq 4n\sum_{i=1}^{\alpha}\binom{c_i}{2} + n\sum_{i=1}^{\alpha}\binom{c_i}{2} \leq 3nC.$$

We consider the following upper bound before invoking Theorem 2. 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.

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

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 2, it holds that

$$\left|\frac{h(\mathcal{G})(2^n-1)^q}{(2^n)_{|\mathcal{V}|}} - 1\right| \le \frac{4S + 14nC + 36v_m}{2^n} + \frac{4Cq_c^2 + 62q_cq_m^2}{2^{2n}} + \frac{40q_m^4}{2^{3n}} =: \epsilon_1(\tau). \quad (7)$$

6.2 Proof of Theorem 7

We will use Theorem 1 to prove the main theorem in this section given Equations (5) and (7). 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. Let $B := 2^n - 1$.

The expectations of the squared terms can be bound as follows. The first expectation is derived from Fact 5. In the second expectation, we use $C = \sum_{i=1}^{\alpha} c_i^2 \le \xi_{\max} \sum c_i \le \xi_{\max} q_c$ and Fact 3 and Fact 5. In the third expectation, we use $Cq_c^2/2^{2n} \le 1$, proven in the previous section.

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

In the last inequality, $q_c \le 2\mu_m + q_x$ and $\mathbf{Ex}[q_x] \le q_m^2\delta$, where q_x is from the proof of Fact 5. We also use $\mu_m \le \frac{2^{0.5n}}{12\sqrt{n}}$ and $q_x \le q_c \le \frac{2^{2n}}{186q_m^2}$ by $\neg\mathsf{bad}_6$.

We derive an upper bound of $\sqrt{2\mathbf{Ex}[\epsilon_1(\tau)^2]}$ using Lemma 5 as follows:

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

where we use $q_m \le \frac{2^{3n/4}}{8}$, $2^{n/8} > n$ and $\delta = \ell/2^n$. Combining it with Equation (5), and using $(4\mu_m^2 + q_m^3\delta)^{0.5} \le c\max(\mu_m, q_m^{1.5}\delta^{0.5})$ for appropriate constant c, the overall bound $\sqrt{2u\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]} + 2u\epsilon_2$ from Theorem 1 is given by

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

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

We balance the last equation by choosing

$$L_1^2 = \frac{3\ell u\mu_m q_m}{\max(157nu^{0.5}\mu_m, 157nu^{0.5}q_m^{1.5}\delta^{0.5}, 3uq_m^2\delta)}, \tag{8}$$

and

$$L_2 = \begin{cases} \left(\frac{3\ell u^{0.5}q_m^2}{157n\max(\mu_m, q_m^{1.5}\delta^{0.5})}\right)^{\frac{1}{3}} & \text{if } q_m^3 < 2^{2n}, \\ \frac{2^n}{32q_m} & \text{if } q_m^3 \ge 2^{2n}. \end{cases} \tag{9}$$

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 $\frac{222\ell uq_m^{2.5}}{2^{2n}} \le \frac{11.1\ell nuq_m^{2.5}}{2^{2n}} \le \left(\frac{11.1\ell nuq_m^{2.5}}{2^{2n}}\right)^{2/3}$ and the AM-GM inequality to suppress some terms as follows:

$$\begin{aligned}
\frac{\ell u\mu_m^2}{2^n} + \frac{3\ell^{\frac{1}{3}}n^{\frac{2}{3}}u^{\frac{2}{3}}\mu_m^{\frac{2}{3}}q_m^{\frac{2}{3}}}{2^n} &\ge \frac{4\ell^{\frac{1}{2}}n^{\frac{1}{2}}u^{\frac{3}{4}}\mu_m q_m^{\frac{1}{2}}}{2^n}, \\
\frac{\ell u\mu_m^2}{2^n} + \frac{3\ell^{\frac{2}{3}}n^{\frac{2}{3}}u^{\frac{2}{3}}q_m^{\frac{5}{3}}}{2^{\frac{4n}{3}}} &\ge \frac{4\ell^{\frac{3}{4}}n^{\frac{1}{2}}u^{\frac{3}{4}}\mu_m^{\frac{1}{2}}q_m^{\frac{5}{4}}}{2^{\frac{5n}{4}}}.
\end{aligned}$$

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 (8) and (9). 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. We first observe that the inequalities $n^2 u\mu_m^2 q_m^3 \delta \ll 2^{2n}$ and $uq_m^4\delta \ll 2^{2n}$ must hold, otherwise the advantage bound becomes larger than 1.

For L_1, choosing $nu^{0.5}q_m^{1.5}\delta^{0.5}$ among three choices for the maximum gives

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

which is much smaller than $\frac{2^n}{n}$ because it is equivalent to $u\mu_m^2\delta \leq q_m$ while we assumed $u\mu_m^2\delta = O(1)$ in the condition. Also, the same choice implies that $L_1 = O(2^n/q_m)$, which is equivalent to $\ell u\mu_m^2 q_m^3 \ll n^2 2^{3n}$, which is true because of the condition $n^2 u\mu_m^2 q_m^3 \delta \ll 2^{2n}$.

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 $\ell n u q_m \ll 2^{2n}$. This is also smaller than $\left(\frac{2^n}{q_m}\right)^3$, which is equivalent to $\ell u q_m^7 \ll n^2 2^{5n}$. This is true because of $u q_m^4 \delta \ll 2^{2n}$ and $q_m^3 < 2^{2n}$. If $q_m^3 \geq 2^{2n}$, $\frac{2^n}{q_m} \ll \frac{2^{0.5n}}{\sqrt{n}}$ apparently holds.

6.3 Further Security Analyses of nEHtM

Further security analyses for nonce-respecting setting, describing problems in [14], proof using a stronger hash, and recovering the result of [14] are in Appendix D.

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 [23] 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 $\mathsf{H}^1, \mathsf{H}^2 : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^{n-1}$. In other words, $\mathsf{H}_{K_h}(M) = (\mathsf{H}^1_{K_{h_1}}(M), \mathsf{H}^2_{K_{h_2}}(M))$ 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:

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

Proving PRF security of deterministic MACs suffices to show their MAC security. Using similar reasoning, one can see proving PRF* security of DbHtS also suffices to show its MAC security.

We also introduce the additional parameter q_m denoting the maximum number of queries each user makes and assume $q = u q_m$ for our security analysis; this does not lose the generality by making some redundant queries at the end.

Theorem 8 shows the multi-user DbHtS security bound improved from [42] (Recall Figure 4 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 is implemented by the ideal cipher; and E is implemented by the ideal cipher. The proof can be found in Appendix E.3.

Theorem 8. *Let n, k, u, p, l, and q_m be positive integers such that $p + u q_m l \leq 2^{n-2}$. 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*

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

Theorem 9 shows an improved result from the previous work [24]. Following the original paper, we require the hash functions H^1, H^2 used in DbHtS to be regular and AU, and is *not* implemented by the ideal cipher; and E is implemented by the ideal cipher. We don't need the assumption that H^1 and H^2 are cross-collision resistant that are originally required in [24]. Instead, we require the hash function is 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:

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

We defer the proof to Appendix E.4.

Theorem 9. *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*

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

Acknowledgement. Wonseok Choi, Yu Wei, and Vassilis Zikas were supported in part by AnalytiXIN, by Sunday Group, and by Algorand Foundation via the MEGA-ACE project. Yu Wei and Vassilis Zikas were supported in part by NSF grant no. 2055599. Wonseok Choi was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.RS-2023-00239620). Minki Hhan was supported by a KIAS Individual Grant QP089801.

References

[1] Bao, Z., Hwang, S., Inoue, A., Lee, B., Lee, J., Minematsu, K.: XOCB: Beyond-birthday-bound secure authenticated encryption mode with rate-one computation. In: EUROCRYPT 2023, Part IV. LNCS, vol. 14007, pp. 532–561 18

[2] 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 6

[3] Bellare, M., Desai, A., Jokipii, E., Rogaway, P.: A concrete security treatment of symmetric encryption. In: 38th FOCS. pp. 394–403 5

[4] 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 5

[5] Bellare, M., Kilian, J., Rogaway, P.: The security of cipher block chaining. In: CRYPTO'94. LNCS, vol. 839, pp. 341–358 5

[6] 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 5, 6

[7] Bellare, M., Namprempre, C.: Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In: ASIACRYPT 2000. LNCS, vol. 1976, pp. 531–545 4

[8] Bellare, M., Namprempre, C.: Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. Journal of Cryptology **21**(4), 469–491 4

[9] 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 4, 5

[10] Bernstein, D.J.: How to stretch random functions: The security of protected counter sums. Journal of Cryptology **12**(3), 185–192 5

[11] 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 6

[12] 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 5

[13] Chen, S., Steinberger, J.P.: Tight security bounds for key-alternating ciphers. In: EUROCRYPT 2014. LNCS, vol. 8441, pp. 327–350 73

[14] Chen, Y.L., Choi, W., Lee, C.: Improved multi-user security using the squared-ratio method. In: CRYPTO 2023, to apear 2, 6, 8, 9, 10, 11, 12, 14, 23, 24, 25, 32, 68, 69, 70, 71

[15] 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 8, 12

[16] 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 6

[17] Choi, W., Lee, B., Lee, J.: Indifferentiability of truncated random permutations. In: ASIACRYPT 2019, Part I. LNCS, vol. 11921, pp. 175–195 6

[18] 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 5, 18

[19] 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 7, 9, 10, 12, 24, 38, 40

[20] Choi, W., Lee, J., Lee, Y.: Building prfs from tprps: Beyond the block and the tweak length bounds. IACR Cryptol. ePrint Arch. p. 918 47

[21] Cogliati, B., Lampe, R., Patarin, J.: The indistinguishability of the XOR of k permutations. In: FSE 2014. LNCS, vol. 8540, pp. 285–302 6

[22] 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 5, 6, 9, 15, 23

[23] 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 7, 8, 9, 32

[24] 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 8, 9, 11, 12, 33, 72, 81, 82, 84, 88

[25] Dutta, A., Nandi, M., Saha, A.: Proof of mirror theory for $\xi_{\max} = 2$. IEEE Trans. Inf. Theory **68**(9), 6218–6232 6

[26] 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 7, 9, 23

[27] 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 6

[28] Gunsing, A., Mennink, B.: The summation-truncation hybrid: Reusing discarded bits for free. In: CRYPTO 2020, Part I. LNCS, vol. 12170, pp. 187–217 6

[29] Guo, T., Wang, P.: A note on the security framework of two-key dbhts macs. In: Information and Communications Security - 24th International Conference, ICICS 2022, Canterbury, UK, September 5-8, 2022, Proceedings. Lecture Notes in Computer Science, vol. 13407, pp. 55–68. https://doi.org/10.1007/978-3-031-15777-6_4 8

[30] Hall, C., Wagner, D., Kelsey, J., Schneier, B.: Building PRFs from PRPs. In: CRYPTO'98. LNCS, vol. 1462, pp. 370–389 5, 6

[31] Impagliazzo, R., Rudich, S.: Limits on the provable consequences of one-way permutations. In: CRYPTO'88. LNCS, vol. 403, pp. 8–26 5

[32] Jha, A., Nandi, M.: Tight security of cascaded LRW2. J. Cryptol. **33**(3), 1272–1317, https://doi.org/10.1007/s00145-020-09347-y 12

[33] 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 8, 9

[34] Landecker, W., Shrimpton, T., Terashima, R.S.: Tweakable blockciphers with beyond birthday-bound security. In: CRYPTO 2012. LNCS, vol. 7417, pp. 14–30 12

[35] 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 6

[36] Lucks, S.: The sum of PRPs is a secure PRF. In: EUROCRYPT 2000. LNCS, vol. 1807, pp. 470–484 6

[37] 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 12

[38] Namprempre, C., Rogaway, P., Shrimpton, T.: Reconsidering generic composition. In: EUROCRYPT 2014. LNCS, vol. 8441, pp. 257–274 4, 9, 10, 12, 17, 18

[39] 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 6

[40] 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 6

[41] 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 6

[42] 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 8, 9, 11, 12, 32, 72, 74, 75, 79

[43] Shrimpton, T.: A characterization of authenticated-encryption as a form of chosen-ciphertext security. Cryptology ePrint Archive 17

Supplementary Material

A Useful Inequalities

We use the following inequalities multiple times in the proofs.

$$\prod_{i=1}^{n}(1-x_i) \geq 1 - \sum_{i=1}^{n} x_i \text{ if } 0 \leq x_i \leq 1 \text{ for all } i \tag{10}$$

$$\sum_{i=1}^{n} i^k \geq \frac{n^{k+1}}{k+1} \text{ for } k \geq 1 \tag{11}$$

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

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

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

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

Lemma 4. *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 $t > 0$:

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

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

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

For the variance, it holds that

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

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

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

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

Lemma 5. *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 $\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]$.

B Proof of Mirror Theory

B.1 Proof of Mirror Theory - Lower Bound for $\xi_{max} > 2$

Below, we describe the proof of Theorem 4, 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 [19] 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 6. *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*

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

In particular, the following inequality holds

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

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 7. *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 $0 \leq i < \alpha$ such that $c_{i+1}C_i \leq N$, it holds that

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

Proof (of claim). By applying Lemma 6 to h_{i+1}, we obtain

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

The second term is bounded below by

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

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

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

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

When $c_{i+1}C_i \leq N$ holds for all $i \leq \alpha - 1$, we obtain the desired result by multiplying the inequalities from the claim for $i = 1, ..., \alpha - 1$ and using Inequality (10), and the fact that $C_i \leq C_\alpha$ for $i \leq \alpha$. □

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 8. *Suppose that $4C_{\alpha+\beta} + 2 \leq N$. Then it holds that*

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

Proof. We use the following claim.

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

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

Proof (of claim). We adapt the inequality bottom of [19, 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 [19, 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. □

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 (10). In the last inequality, we use Inequality (11) and $\beta \leq N$. □

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 9. *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 $C_{\alpha+\beta} + |\mathcal{D}| \leq N/2$. □

Proof of Theorem 4. Observe that $8q \leq N$ implies the conditions of all lemmas. Applying Lemmas 7 to 9 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. □

B.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 5.

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

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

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 $\mathcal{G}_q = \mathcal{G}(\mathit{\Gamma})$, which is the graph representation of the equation system $\mathit{\Gamma}$. We also define

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

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) $\mathcal{F}[\mathcal{J}, \mathcal{L}] \overset{\text{def}}{=}$

$\{(P_{\gamma_{j_1}}, P_{\gamma'_{j_2}}, L_1), \cdots, (P_{\gamma_{j_k}}, P_{\gamma'_{j_{k+1}}}, L_k)\}$ and a weighted graph (an equation system) $\mathcal{G}[\mathcal{I}, \mathcal{J}, \mathcal{L}] \overset{\text{def}}{=} \mathcal{G}[\mathcal{I}] \cup \mathcal{F}[\mathcal{J}, \mathcal{L}]$. 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

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

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

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 $\Lambda_t \overset{\text{def}}{=} (\bigsqcup_{i \in [t]} \mathcal{C}_i)$. We thus have

$$\begin{aligned} h_t &= \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} \Big(N - |\Lambda_{t-1} \bigcup (\Lambda_{t-1} \oplus \lambda_t)| \Big) \\ &= \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} \Big(N - |\Lambda_{t-1}| - |\Lambda_{t-1} \oplus \lambda_t| + |\Lambda_{t-1} \bigcap (\Lambda_{t-1} \oplus \lambda_t)| \Big) \\ &= (N - 2C_{t-1})h_{t-1} + \sum_{(V_1, \ldots, V_{t-1}) \in \mathcal{S}_{t-1}} |\Lambda_{t-1} \bigcap (\Lambda_{t-1} \oplus \lambda_t)|, \end{aligned} \tag{12}$$

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

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

This leads to

$$\begin{aligned}
&\sum_{(V_1,\ldots,V_{t-1}) \in \mathcal{S}_{t-1}} \sum_{V,V' \in \Lambda_{t-1}} \mathbb{1}(V \oplus V' = \lambda_t) \\
= &\sum_{(V_1,\ldots,V_{t-1}) \in \mathcal{S}_{t-1}} \left(\sum_{(W,W',V,V') \in \mathcal{R}_t} \mathbb{1}(V \oplus V' = \lambda_t) + \sum_{V \in C_i, V' \in C_j, i<j<t} \mathbb{1}(V \oplus V' = \lambda_t) \right) \\
= &\ |\mathcal{R}_t|\, h_{t-1} + \sum_{E \in \mathcal{L}[\mathcal{G}_t]} h(\mathcal{G}_{t-1} \cup E). \qquad (13)
\end{aligned}$$

Lemma 10 follows from Equations (12) and (13). □

Lemma 11 (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

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

When $P_{\gamma_{m-1}} \in (\mathit{\Lambda}_{m-1}\oplus L_k\oplus\lambda_m)$, we know that the vertex $V = P_{\gamma_m}$ must satisfy $V \in \mathit{\Lambda}_{m-1}\setminus\mathcal{V}[\mathcal{J}]$. 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 $P_{\gamma_{m-1}} \in (\mathit{\Lambda}_{m-1}\oplus L_k)$, we know there exists a vertex $V = P_{\gamma_m}$ must satisfy $V \in \mathit{\Lambda}_{m-1}\setminus\mathcal{V}[\mathcal{J}]$, 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

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

where

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

When $P_{\gamma_{m-1}} \in (\mathit{\Lambda}_{m-1}\oplus L_k)\cap(\mathit{\Lambda}_{m-1}\oplus L_k\oplus\lambda_m)$, we know there exists two distinct vertecies $V', W \in \mathit{\Lambda}_{m-1}\setminus\mathcal{V}[\mathcal{J}]$ 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

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

where

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

This conclude the proof. □

Lemma 12 estimates the size of sets $\mathbb{L}[\mathcal{G}_m]$, $\mathbb{M}[\mathcal{G}^{++}]$, and $\mathbb{N}[\mathcal{G}^{++}]$ using in Lemma 10 and 11. In order to state Lemma 12, 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

$$\max_{\lambda \in \{0,1\}^n \setminus \{0^n\}} \{|\{k \leq m : \lambda_k = \lambda\}|\} \leq |\mathcal{R}_{m+1}|\,. \tag{14}$$

Lemma 12 (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

$$(m - 1 - |\mathcal{R}_m|)(m - 1 - |\mathcal{R}_m| - 1), \tag{15}$$

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 14,

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

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

$$\left|\mathbb{M}[\mathcal{G}^{-+}]\right| - 4(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{M}[\mathcal{G}^{++}]\right| \leq 2r. \tag{16}$$

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 14

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

   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

$$\left|\mathbb{N}[\mathcal{G}^{-+}]\right| - 4r(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{N}[\mathcal{G}^{++}]\right| \leq r^2. \tag{17}$$

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

$$2m - |\mathcal{R}[\mathcal{I}]_m| - 4(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{M}[\mathcal{G}^{++}]\right| \leq 2r. \tag{18}$$

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
$$\left|\mathbb{L}[\mathcal{G}^{-+}]\right| - 4r(|\mathcal{R}_{t+1}| + 1) \leq \left|\mathbb{N}[\mathcal{G}^{++}]\right| \leq r^2. \tag{19}$$

By Equation (15) and inequalities (16) to (19), the proof is completed. □

The following combinatorial lemma proved by [20] is used in our Mirror theory statement.

Lemma 13. *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 14. *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 $V \in \Lambda_m$, $P_{\gamma_{m+1}} \neq V, P_{\gamma'_{m+1}} \neq V$. Therefore, we have

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

By repeatedly applying the above inequality, we have

$$h(\mathcal{G}_m) \leq \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m}}, \tag{20}$$

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 $V \in \Lambda_m \setminus \mathcal{C}_{m-k}$, $P_{\gamma_{m-k}} \neq V, P_{\gamma'_{m-k}} \neq V$. Therefore, we have

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

By repeatedly applying the above inequality, we have

$$h(\mathcal{G}^{++}) \leq \frac{h(\mathcal{G}_m)}{(N-2r)^k}. \tag{21}$$

Combining Equation (20) and (21), we complete the proof. □

By lemma 14, 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,

$$D^t_{m,k} \leq \max\{\frac{h(\mathcal{G}^{-+})}{N}, h(\mathcal{G}^{++})\} \leq \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+k}}. \tag{22}$$

Lemma 15 shows that our constructed two-dimensional sequence $D^t_{m,k}$ satisfies the condition required for using combinatorial Lemma 13. This proof is based on purple equation (Lemma 11), size lemma (Lemma 12) and Lemma 14.

Lemma 15. *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

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

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 11), we have

$$h(\mathcal{G}^{++}) = h(\mathcal{G}^{+-}) - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\}) + \sum_{\{E,E'\} \in \mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E, E'\}), \tag{23}$$

$$h(\mathcal{G}^{-+}) = h(\mathcal{G}^{--}) - \sum_{E \in \mathbb{M}[\mathcal{G}^{-+}]} h(\mathcal{G}^{--} \cup \{E\}) + \sum_{\{E,E'\} \in \mathbb{N}[\mathcal{G}^{-+}]} h(\mathcal{G}^{--} \cup \{E, E'\}). \tag{24}$$

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

$$\begin{aligned}
&\left| \sum_{E \in \mathbb{M}[\mathcal{G}^{-+}]} \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\}) \right| \\
&\leq \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} \left| \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} - h(\mathcal{G}^{+-} \cup \{E\}) \right| + \sum_{E \in \mathbb{M}[\mathcal{G}^{-+}] \setminus \mathbb{M}[\mathcal{G}^{++}]} \left| \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} \right| \\
&\leq 2r \cdot D^t_{m-1,k} + 4(|\mathcal{R}_{t+1}| + 1) \left| \frac{h(\mathcal{G}^{--} \cup \{E\})}{N} \right| \quad \text{(by size lemma (Lemma 12))} \\
&\leq 2r \cdot D^t_{m-1,k} + \frac{4(|\mathcal{R}_{t+1}| + 1) h(\mathcal{G}_t)}{N(N-2r)^{t-m+k}}. \quad \text{(by Lemma 14)}
\end{aligned}$$

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 12 and 14, 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 (23) from $\frac{1}{N}\times$ Equation (24) and combine everything above, we have

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

When $m \geq 4$ 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 11) and Orange equation (Lemma 10), respectively, we have

$$h(\mathcal{G}^{++}) = h(\mathcal{G}^{+-}) - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\}) + \sum_{\{E,E'\} \in \mathbb{N}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E, E'\}), \tag{25}$$

$$h(\mathcal{G}^{-+}) = h(\mathcal{G}^{--}) - (2m-2-|\mathcal{R}[\mathcal{I}]_m|)h(\mathcal{G}^{--}) + \sum_{E \in \mathcal{L}[\mathcal{G}^{-+}]} h(\mathcal{G}^{--} \cup E). \tag{26}$$

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

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

Using the above inequality, we have

$$\begin{aligned}
&\left|(2m-2-|\mathcal{R}[\mathcal{I}]_m|)\frac{h(\mathcal{G}^{--})}{N} - \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} h(\mathcal{G}^{+-} \cup \{E\})\right| \\
&\leq \sum_{E \in \mathbb{M}[\mathcal{G}^{++}]} \left|\frac{h(\mathcal{G}^{--})}{N} - h(\mathcal{G}^{+-} \cup \{E\})\right| + \left|2m-2-|\mathcal{R}[\mathcal{I}]_m| - \left|\mathbb{M}[\mathcal{G}^{++}]\right|\right| \frac{h(\mathcal{G}^{--})}{N} \\
&\leq 2r \cdot D^t_{m-1,1} + 4(|\mathcal{R}_{t+1}|+1)\frac{h(\mathcal{G}^{--})}{N} \qquad\qquad\qquad\qquad \text{(by Lemma 12)} \\
&\leq 2r \cdot D^t_{m-1,1} + \frac{4(|\mathcal{R}_{t+1}|+1)h(\mathcal{G}_t)}{N(N-2r)^{t-m+1}}.
\end{aligned}$$

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

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

It implies that

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

By subtracting Equation (25) from $\frac{1}{N}\times$ Equation (26) and combining the above, we have

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

This concludes the proof. □

When $k=1$, Lemma 16 gives a sharper upper bound on $D^t_{t,1}$. The proof can be derived from Lemma 15 and 13.

Lemma 16. *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 15, let $n \le \frac{m}{2} - 1$ (the c in Lemma 13), then we can apply Lemma 13 to obtain

$$\begin{aligned}
D^t_{m,1} &\le \sum_{i=n}^{2n} \binom{2n}{i} r^i D^t_{m-n,1-n+i} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j} \binom{2j}{i} \frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\le \sum_{i=n}^{2n} (2e)^i r^i D^t_{m-n,1-n+i} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j} \binom{2j}{i} \frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\qquad\qquad (\tbinom{2n}{i} \le (\tfrac{2ne}{i})^i \le (2e)^i \text{ when } n \le i \le 2n) \\
&\le \sum_{i=n}^{2n} (2er)^i \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+1+i}} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j} \binom{2j}{i} \frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\qquad\qquad \text{(by Inequality 22)} \\
&= \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \sum_{i=n}^{2n} (\frac{2er}{N-2r})^i + \sum_{j=0}^{n-1}\sum_{i=j}^{2j} \binom{2j}{i} \frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\le \frac{h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \sum_{i=n}^{\infty} (1/2)^i + \sum_{j=0}^{n-1}\sum_{i=j}^{2j} \binom{2j}{i} \frac{r^i C}{(N-2r)^{t-m+1+i}} \quad (r \le \tfrac{N}{13}) \\
&\le \frac{2h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \frac{1}{2^n} + \sum_{j=0}^{n-1}\sum_{i=j}^{2j} \binom{2j}{i} \frac{r^i C}{(N-2r)^{t-m+1+i}} \\
&\le \frac{2h(\mathcal{G}_t)}{(N-2r)^{t-m+1}} \frac{1}{2^n} + \frac{4C}{(N-2r)^{t-m+1}}.
\end{aligned}$$

Now, plug in $m = t$ and have

$$\begin{aligned}
D^t_{t,1} &\le \frac{2h(\mathcal{G}_t)}{(N-2r)} \frac{1}{2^n} + \frac{4C}{(N-2r)} \\
&\le \frac{\frac{26}{11} h(\mathcal{G}_t)}{N^2} + \frac{\frac{13}{11}(24\,|\mathcal{R}_{t+1}| + 24) h(\mathcal{G}_t)}{N^2} \\
&\qquad (\because r \le \tfrac{N}{13}, \text{ substitute } C \text{ defined in Lemma 15}) \\
&= \frac{(29\,|\mathcal{R}_{t+1}| + 31) h(\mathcal{G}_t)}{N^2}.
\end{aligned}$$

This concludes the proof. □

Finally, using the above, we can prove Theorem 5 as follows:

Proof (of Theorem 5). 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 6, we simply have $(N - c_{i+1}C_i)h_i \leq h_{i+1}$ for $i = 0, \cdots, 2n+1$ as $v_{i+1} = 0$ in graph $\mathcal{G}$, which represents the equation system $\varGamma$.

For $i \geq 2n+2$, we first replicate part of the proof of Lemma 10 and have

$$h_{i+1} = (N - 2C_i)h_i + |\mathcal{R}_{i+1}|h_i + \sum_{\{V,V'\} \in \mathbb{L}_{i+1}} h'(V, V'), \tag{27}$$

where recall $h'(V, V')$ denote the number of solutions to $\varLambda_i$ such that $V \oplus V' = \lambda_{i+1}$ for $V, V' \in \varLambda_i$, and $\mathbb{L}_{i+1} \overset{\text{def}}{=} \big\{\{V, V'\} \in \varLambda_i^{*2} \,\big|\, \lambda(V, V') = \perp\big\}$. We also have $|\mathbb{L}_{i+1}| = C_i(C_i - 2) = 4i^2 - 4i$. Then by Lemma 16, we have

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

Plugging in Equation (27), we have

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

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

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

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

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

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. □

B.3 Proof of Mirror Theory - Upper Bound for $\xi_{\max} > 2$

Theorem 10 (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

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

The proof of Theorem 10 is deferred to the end of this section. Before proving it, we introduce essential lemmas first.

Lemma 17. *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 $\Lambda_V = (\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i) \oplus \lambda_{i+1}(V)$. 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

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

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

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

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

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

Then the summation $\sum_{(V_1,\ldots,V_i)\in\mathcal{S}_i}\sum_{V,V'\in C_{i+1}}|\varLambda_V\cap\varLambda_{V'}|$ 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 $\varLambda_{V_1} \cap \varLambda_{V_1'}$ 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:

$$
\begin{aligned}
h'(V_1, V_1', V_2, V_2') &\le N \cdot h''(V_2, V_2') && \text{(Upper bound of Lemma 6)}\\
&\le \frac{N h_i}{(N - \xi_{\max} C_i)^2} && \text{(Lower bound of Lemma 6)}\\
&\le \frac{h_i}{N}\left(1 + \frac{2N\xi_{\max}C_i}{(N - \xi_{\max}C_i)^2}\right)\\
&\le \frac{h_i}{N}\left(1 + \frac{192\xi_{\max}q_c}{25N}\right)\\
&\le \frac{73h_i}{25N},
\end{aligned}
$$

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. □

Lemma 18. *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 $\Lambda_i \overset{\text{def}}{=} \bigsqcup_{j \in [i]} \mathcal{C}_i$ for $i = \alpha, \cdots, \alpha + \beta - 1$. We thus have

$$\begin{aligned} h_{i+1} &= \sum_{(V_1, \ldots, V_i) \in \mathcal{S}_i} \left(N - |\Lambda_i \bigcup (\Lambda_i \oplus \lambda_{i+1})| \right) \\ &= \sum_{(V_1, \ldots, V_i) \in \mathcal{S}_i} \left(N - |\Lambda_i| - |\Lambda_i \oplus \lambda_{i+1}| + |\Lambda_i \bigcap (\Lambda_i \oplus \lambda_{i+1})| \right) \\ &= (N - 2C_i)h_i + \sum_{(V_1, \ldots, V_i) \in \mathcal{S}_i} |\Lambda_i \bigcap (\Lambda_i \oplus \lambda_{i+1})|. \end{aligned} \tag{28}$$

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

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

Then we have

$$\sum_{(V_1, \ldots, V_i) \in \mathcal{S}_i} |\Lambda_i \bigcap (\Lambda_i \oplus \lambda_{i+1})| = |\mathcal{R}_{i+1}| h_i + \sum_{\{V, V'\} \in \mathbb{M}_{i+1}} h'(V, V'). \tag{29}$$

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

$$\frac{(c_j C_i)^2}{N - c_j C_i} \leq c_j C_i, \text{ and } \frac{(c_j C_i)^2}{N - c_j C_i} \leq \frac{2(c_j C_i)^2}{N}.$$

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

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

where we use $\sum_{i=1}^{\alpha} c_i = C_\alpha \leq 3q_c/2$, $C_i \leq q$, $i \leq \beta \leq q$ and $c_j = 2$ for all $j \geq \alpha + 1$ 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 $i \geq 2n+2$, by Lemma 16, then we have

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

equivalently, we have

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

Plugging in Equation (29), we have

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

Plugging the above inequality into Equation (28), we have

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

This concludes the proof. □

Using the above lemmas, we can prove Theorem 10 as follows.

Proof (of Theorem 10). We start by finding a relation between h_i and h_{i+1}. Lemma 6 has already shown $h_{i+1} \leq N h_i$, while Lemmas 17 and 18 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 $\xi_{\max} \geq 3$, hence $q \leq \frac{N}{12}$ by the constraints $4q\xi_{\max} \leq N$. 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

$$(N - C_i)_{c_{i+1}} \geq N^{c_{i+1}-1}(N - c_{i+1}C_{i+1}), \tag{30}$$

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

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

So we have

$$\begin{aligned}
&\frac{h_{i+1}(N-1)^{c_{i+1}-1}}{h_i(N-C_i)_{c_{i+1}}} \\
&\leq \frac{(N-1)^{c_{i+1}-1}\left(N - c_{i+1}C_i + |\mathcal{R}_{i+1}| + \frac{2(c_{i+1})_2 q_c^2}{N}\right)}{N^{c_{i+1}-1}(N - c_{i+1}C_{i+1})} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{(by Lemma 17 and Equation (30))} \\
&\leq 1 + \frac{c_{i+1}C_{i+1} - c_{i+1}C_i}{N - c_{i+1}C_{i+1}} + \frac{|\mathcal{R}_{i+1}|}{N - c_{i+1}C_{i+1}} + \frac{2(c_{i+1})_2 q_c^2}{N(N - c_{i+1}C_{i+1})} \\
&\leq 1 + \frac{c_{i+1}^2}{N - c_{i+1}C_{i+1}} + \frac{|\mathcal{R}_{i+1}|}{N - c_{i+1}C_{i+1}} + \frac{2(c_{i+1})_2 q_c^2}{N(N - c_{i+1}C_{i+1})} \\
&\leq 1 + \frac{2c_{i+1}^2}{N} + \frac{2|\mathcal{R}_{i+1}|}{N} + \frac{4(c_{i+1})_2 q_c^2}{N^2}. \qquad (c_{i+1}C_{i+1} \leq \tfrac{N}{2} \text{ by } 4q\xi_{\max} \leq N)
\end{aligned}$$

Now we can compute

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

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

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

Now we can compute

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

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 $1+x \leq e^x$, $\beta \leq q$, 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. □

B.4 Proof of Mirror Theory - Upper Bound for $\xi_{\max} = 2$

Theorem 11 (Upper Bound Mirror Theory). *Let $\mathcal{G}$ be a nice graph and q denote the number of edges of $\mathcal{G}$. When $q \leq \frac{N}{13}$ and $q_c = 0$, it holds that*

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

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

Lemma 19. *For $i \in [2n+2, \beta-1]$, it holds that*

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

Proof. For $V, V' \in \mathbb{L}_{i+1}$, if $i \geq 2n+2$, by Lemma 16, then we have

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

equivalently, we have

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

Plugging in Equation (29), we have

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

Plugging the above inequality into Equation (28), we have

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

This completes the proof. □

Using Lemma 19, we can prove Theorem 11 as follows.

Proof (of Theorem 11). 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 $i = 2n+2, \cdots, \beta-1$. For $i = 2n+2, \cdots, \beta-1$, using Lemma 19, we have

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

By using the above inequality, we have

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

which completes the proof. □

C Proof of Multi-User Security of XoP

We will use the squared-ratio method to prove an upper bound for $\mathsf{Adv}_{\mathsf{XoP}}^{\mathsf{mu\text{-}prf}^*}(u, q_m)$, where on each of u user the adversary can make at most q_m queries. To prove this, it suffices to consider an information-theoretic adversary $\mathcal{D}$ making q_m queries on a single user. Denote the the queries made by $\mathcal{D}$ as $x_1, ..., x_{q_m} \in \{0,1\}^{n-1}$, and denote the transcript as $\tau = ((x_1, \lambda_i), \cdots, (x_{q_m}, \lambda_{q_m}))$. Without loss of the generality, we assume all queries are different.

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 $\{P_{\gamma_1}, P_{\gamma_1'}, \cdots, P_{\gamma_n}, P_{\gamma_n'}\}$ is a solution of the following equation system

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

The equations system induces a transcript graph $\mathcal{G}(\tau)$ in the real world.

Bad transcript analysis. Recall in the ideal world, for every query $\mathsf{XoP}[\mathsf{P}](x_i)$, the query result λ_i is taken uniformly at random from $\{0,1\}^n \setminus \{\mathbf{0}\}$. We define the bad event bad in the ideal world as n different queries having the same query output:

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

Let T_{re} be a random variable following the distribution of the transcripts in the ideal world. We have

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

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

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

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

The expression holds is because, in the real world, the transcript occurs with a probability proportional to the number of solutions $h(\mathcal{G})$ over all possible choices of $P_{\gamma_i}, P_{\gamma_i'}$. In the ideal world, every transcript occurs equally, i.e., with probability $\frac{1}{(2^n-1)^{q_m}}$.

Recall the definition of $\mathcal{R}_i$,

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

For any good transcript τ, we have $|\mathcal{R}_i| \leq n$ for all $i \in [q_m]$ because of $\neg\mathsf{bad}$.

Let $I_{j,k}$ be the indicator random variable that equals to 1 if $\lambda_j = \lambda_k$ and 0 otherwise. Then it holds that $\sum_{i=1}^{q_m} |\mathcal{R}_i| = \sum_{j,k \in [q_m]} I_{j,k}$. Given $\neg\mathsf{bad}$, we have

$$\frac{3\sum_{i=1}^{q_m} |\mathcal{R}_i|}{2^n} + \frac{12q_m^2}{2^{2n}} + \frac{10(n+1)^2}{2^n} \leq \frac{3nq_m}{2^n} + \frac{12q_m^2}{2^{2n}} + \frac{10(n+1)^2}{2^n} \leq 1$$

for $n > 12$ and $q_m \leq \frac{2^n}{4n}$. Therefore, by Theorem 3,

$$\left|\frac{h(\mathcal{G})(2^n-1)^{q_m}}{(2^n)_{2q_m}} - 1\right| \leq \frac{6\sum_{i=1}^{q_m} |\mathcal{R}_i|}{2^n} + \frac{24{q_m}^2}{2^{2n}} + \frac{20(n+1)^2}{2^n}.$$

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

$$\epsilon_1(\tau) = \frac{6\sum_{i=1}^{q_m} |\mathcal{R}_i|}{2^n} + \frac{24{q_m}^2}{2^{2n}} + \frac{20(n+1)^2}{2^n}.$$

To apply Theorem 1, we need to bound the expectation of $\epsilon_1(\tau)^2$ where the randomness is taken over the distribution of transcript in the ideal world. Applying Lemma 5 and Lemma 4, 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 1 and plugging in the above inequality, we have

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

This completes the proof. □

D Further Security Analyses of nEHtM

D.1 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 $k \geq 1$ and uniform and independent random variables $T_1, ..., T_k$ sampled from $\{0,1\}^n \setminus \{\mathbf{0}\}$, it holds that for any $K \in \{0,1\}^n$

$$\Pr\left[\bigoplus_{i\in[k]} T_i = K\right] \leq \frac{1}{2^n - 1} = \frac{1}{B}. \qquad (31)$$

We now analyze the probability that each bad event occurs. We assume that $n \geq 20$, and $\mu_m, L_1, L_2 \geq 1$. 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 $k(\neq i) \in [q_m]$ 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

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

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

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

Using $\delta = \ell/2^n$ for $\ell \ge 1$, $B = 2^n - 1 \ge 1.0001 \cdot 2^n$, and $q_m \le 0.01 \cdot 2^n, 2^{3n/4}/8$, 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, $d_X \le L_2$ 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 $\sum_{d_X \ge 2} \binom{d_X}{2} \le L_2^2/2$. 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 $\mathbf{Ex}\left[I_i\right] \leq q_m\delta$ 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

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

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

D.2 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 1, 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 $L_2 = \Theta\left(\min\left(\left(\frac{\ell u q_m 2^n}{n^2}\right)^{\frac{1}{6}}, \frac{2^n}{q_m}\right)\right)$, we have

$$O\left(\frac{\ell u v_m}{2^n} + \left(\frac{\ell u q_m^4}{2^{3n}}\right)^{\frac{1}{2}} + \left(\frac{\ell^2 n^2 u^2 q_m^5}{2^{4n}}\right)^{\frac{1}{3}}\right), \tag{32}$$

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

D.3 The security bound of nEHtM2 in [14]

We briefly sketch the problems in the original multi-user nEHtM2 security proof in [14], 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. [14] 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 [14, page 28], there is a term about $\frac{\sqrt{u} n L q_{\max} \delta}{2^n}$ at the second line. However, there is no corresponding term in the final bound. A similar term $\frac{\sqrt{u} n L q_{\max}^2 \delta}{2^n}$ 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 Appendix D.4 for a more detailed analysis.

D.4 Using Stronger Hash and Proofs in [14]

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 [14] 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 Appendix D.1). 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 $\mathbf{Ex}\left[q_x^2\right] = \mathbf{Ex}\left[\sum_{i,j} I_i I_j\right]$ by

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

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 $\{0,1\}^n\setminus\{\mathbf{0}\}$:

$$\tilde{O}\left(\frac{\sqrt{u}q_m^4}{2^{3n}}+\left(\frac{uq_m^2}{2^{2n}}\right)^{2/3}+\left(\frac{u^2q_m^6}{2^{5n}}\right)^{1/3}\right). \tag{33}$$

Recovering the result of [14] using δ-AXU$^{(2)}$. We give the correct asymptotic bound of the multi-use nEHtM2 security following the proof of [14] assuming that H is δ-AXU$^{(2)}$. Recall the following bound from [14, eprint version, page 27], which are based on the slightly worse version of Theorem 2 with a different bad event for bounding q_c.

$$\begin{aligned}\epsilon_1(\tau)&=\tilde{O}\left(\frac{\sum_{i=1}^{\alpha+\beta}|\mathcal{S}_i|}{2^n}+\frac{Lq_c}{2^n}+\frac{Lq_cq_m^2}{2^{2n}}+\frac{Lq_m^4}{2^{3n}}\right),\\\epsilon_2&=\tilde{O}\left(\frac{q_m^2}{2^{2n}}+\frac{q_m^2}{L^2 2^n}+\frac{L^2q_m^8}{2^{6n}}\right),\end{aligned}$$

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 $L \geq 1$. We stress that the red term in the above bound is missing in the final bound of [14, 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

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

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 $\left(\frac{u^2q_m^6}{2^{5n}}\right)^{1/3}$ in the original bound is just hidden by the other terms; i.e., our analysis does not miss the term.

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

Finally, our new bound in Equation (33) with the same assumption is always tighter than this bound, because

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

Recovering the result of [14] 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, $\Pr[\mathsf{bad}_6] = \tilde{O}\left(\frac{L^2q_m^7}{2^{5n}}.\right)$

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

$$\mathbf{Ex}\left[\epsilon_1(\tau)^2\right]^{1/2} = \tilde{O}\left(\frac{Lq_m^{1.5}}{2^{1.5n}} + \frac{Lq_m^4}{2^{3n}}\right), \quad \epsilon_2 = \tilde{O}\left(\frac{q_m^2}{2^{2n}} + \frac{q_m^2}{L^22^n} + \Pr[\mathsf{bad}_6]\right),$$

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

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

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

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

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

E Proof of Multi-User Security of DbHtS

E.1 More Pictorial Comparison between Security Bound

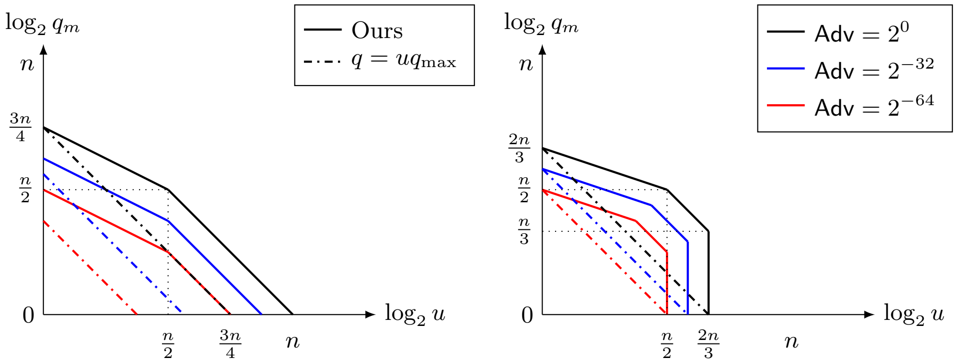

Fig. 4: Comparison of the DbHtS's security bounds (in terms of the threshold number of queries per user) as functions of $\log_2 u$. The black line represents the bound where adversary advantage $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{DbHtS}}(u, q_m, p) = 2^0$; the blue line is for $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{DbHtS}}(u, q_m, p) = 2^{-32}$; and the red line is for $\mathsf{Adv}^{\mathsf{mu-mac}}_{\mathsf{DbHtS}}(u, q_m, p) = 2^{-64}$. 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 9 with Theorem 1 from [24]. The right figure compares our Theorem 8 with Theorem 1 from [42], where we set $\epsilon_3, \epsilon_4 = 0$. We set $p = q_m, k = n, \delta = 2^{-n}$, and neglect l and the logarithmic term of n in all graphs.

Figure 4 gives one more pictorial comparison between our multi-user security bound for DbHtS and previous results by setting the number of primitive queries $p = q_m$, where q_m is the maximum number of construction queries per user

(DbHtS construction instance). Figure 4 shows that the effect of fine-tuning also appears in the low end, for example, when $q_m \le \frac{2n}{3}$ still allows the $\frac{n}{2}$ security bound in both cases when the other parameters are sufficiently small, which was impossible in the previous bounds.

E.2 Useful results for DbHtS proof

We first recall the well-known Patarin's H-coefficient technique, which will be used in both the proof of Theorem 8 and Theorem 9. Recall $\mathcal{S}_1^i$ and $\mathcal{S}_0^i$ are random system of the real world and the ideal world for an i-th user, separately. A transcript z is *attainable* if the probability of obtaining z from $\mathcal{S}_0^i$ is non-zero. The Patarin's H-coefficient technique can be expressed as below:

Lemma 20 ([13]). *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* $\epsilon_1, \epsilon_2 \ge 0$ *be two constants. If* $\frac{\mathsf{p}_{\mathcal{S}_1^1}(z)}{\mathsf{p}_{\mathcal{S}_0^1}(z)} \ge 1 - \epsilon_1$ *holds for all attainable* $z \in \Gamma_{\mathsf{good}}$ *and* $\Pr\left[Z_{\mathcal{S}_0^1} \in \Gamma_{\mathsf{bad}}\right] \le \epsilon_2$*, then, it holds that*

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

E.3 Proof of Theorem 8

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 message M_a^i and let l be the maximal number of blocks of any message M among all 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 $\Sigma_a^i = \mathsf{H}^1_{K_{h,1}}(M_a^i)$ and $\Psi_a^i = \mathsf{H}^2_{K_{h,2}}(M_a^i)$, 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 $U_a^i = \mathsf{E}(K_i, 0\|\Sigma_a^i)$ and $Q_a^i = \mathsf{E}(K_i, 1\|\Psi_a^i)$, 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 $\mathcal{D}$ finishes all its queries, we will further give the following information to the distinguisher $\mathcal{D}$:

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 construction query a.

In the ideal world, we will instead give the distinguisher $\mathcal{D}$:

1. $(K_h^i, K_i) \leftarrow_\$ \{0,1\}^{2k} \times \{0,1\}^k$, independent of its queries.
2. the dummy values U_a^i and Q_a^i computed by the simulation oracle SIM($\mathbb{Q}(J)$) given in Algorithm 1, which is exactly the same as [42, Fig.4].

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.

Algorithm 1 Simulation oracle Sim($\mathbb{Q}(J)$) in the ideal world. For each J, ϕ_J is a partial function that used to simulate a random permutation. The domain and range of ϕ_J are initialized to be the domain and range of E_J respectively.

```
$\forall(\text{eval}, i, M_a^i, T_a^i) \in \mathbb{Q}(J) : (\varSigma_a^i, \varPsi_a^i) \leftarrow \mathsf{H}_{K_h}(M_a^i)$
$I(J) = \{(i,a) : 1 \le i \le u, 1 \le a \le q_m, (\text{eval}, i, M_a^i, T_a^i) \in \mathbb{Q}(J)\}$
$H(J) = \{(\varSigma_a^i, \varPsi_a^i) : (i,a) \in I(J)\}$
$F(J) = \{(i,a) :$ both $\varSigma_a^i$ and $\varPsi_a^i$ are fresh in $H(J)\}$
$G(J) = \{(i,a) :$ only one of $\varSigma_a^i$ or $\varPsi_a^i$ is fresh in $H(J)\}$
$R(J) = \{(i,a) :$ neither $\varSigma_a^i$ nor $\varPsi_a^i$ is fresh in $H(J)\}$
$O(J)$ : set of tuples of $2|F(J)|$ distinct values from $\{0,1\}^n \backslash Rng(\phi_J)$
$S(J) = \{(W_a^i, X_a^i)_{(i,a)\in F(J)} \in O(J) : W_a^i \oplus X_a^i = T_a^i\}$
$(U_a^i, Q_a^i)_{(i,a)\in F(J)} \leftarrow_\$ S(J)$
$\forall(i,a) \in F(J) : (\phi_J(\varSigma_a^i), \phi_J(\varPsi_a^i)) \leftarrow (U_a^i, Q_a^i)$
$\forall(i,a) \in G(J) :$
    if $\varSigma_a^i$ is not fresh in $H$ then
        if $\varSigma_a^i \notin \text{Dom}(\phi_J)$ then $U_a^i \leftarrow_\$ \{0,1\}^n \backslash \text{Rng}(\phi_J); \phi_J(\varSigma_a^i) \leftarrow U_a^i$
        else $U_a^i \leftarrow \phi_J(\varSigma_a^i)$
        $Q_a^i \leftarrow T_a^i \oplus U_a^i$
    else do the same thing to $\varPsi_a^i$ as that to $\varSigma_a^i$
$\forall(i,a) \in R(J) : (U_a^i, Q_a^i) \leftarrow (\bot, \bot)$; **abort**
return $(U_a^i, Q_a^i)_{(i,a)\in I(J)}$
```

Bad Transcript Analysis and Interpretations. We now define bad transcripts. Let T_{id} and T_{re} be random variables following the distribution of a transcript 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* if it is not bad. We refer [42, 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 [42, $\mathsf{bad}_{15}, \mathsf{bad}_{16}$] are excluded by the fine-tuned ideal world (bad_{15}) and the domain separation (bad_{16}) in the analysis below. Instead, to use the mirror theory (Theorem 5) tailored for the fine-tuned ideal world, we consider a new event bad_{15}.

1. bad_1 is the event that there exists a user i such that its left hash key or right hash key equals to its block-cipher key. In other words, $\exists i$ such that $K_i = K^i_{h,1}$ or $K_i = K^i_{h,2}$.
2. bad_2 is the event that there exists a user i both its block cipher key K_i, and left and/or right hash key ($K^i_{h,1}$, $K^i_{h,2}$) has been used by another user j or another primitive query $(\mathsf{prim}, J, x, y, *)$.

Note that $\neg\mathsf{bad}_1 \wedge \neg\mathsf{bad}_2$ guarantees that any user i has at least one fresh key (either hash key or block cipher key). Further, $\neg\mathsf{bad}_3$ guarantees that the distinguisher $\mathcal{D}$ cannot control hash output in the construction query by issuing primitive queries.

3. bad_3 is the event that there exists a construction query $(\mathsf{eval}, i, M^i_a, T^i_a)$ and a primitive query $(\mathsf{prim}, J, x, y, +)$ such that J is used as either left and/or right hash key in the construction query; and one of the message block used in the construction query is queried by the primitive query.

$\neg\mathsf{bad}_4 \wedge \neg\mathsf{bad}_5$ guarantees that neither the inputs nor outputs of the block cipher in construction queries collide with those in the primitive queries when the block cipher key is the same as that of the primitive query.

4. bad_4 is the event that there exists a construction query $(\mathsf{eval}, i, M^i_a, T^i_a)$ and a primitive query $(\mathsf{prim}, J, x, y, +)$ such that the block cipher key in the construction query is J; and at least one of the block cipher's input is x.
5. bad_5 is the event that there exists a construction query $(\mathsf{eval}, i, M^i_a, T^i_a)$ and a primitive query $(\mathsf{prim}, J, x, y, +)$ such that the block cipher key in the construction query is J; and at least one of the block cipher's output is y.

$\neg\mathsf{bad}_6 \wedge \neg\mathsf{bad}_7$ guarantees that if there exists user i and j who share the same block cipher key, then in their construction queries, the inputs of the block cipher are distinct.

6. bad_6 is the event that there exists construction query $(\mathsf{eval}, i, M^i_a, T^i_a)$ and $(\mathsf{eval}, j, M^j_b, T^j_b)$ such that their block cipher keys are the same, namely, $K_i = K_j$; and their left input of the block cipher are the same, namely, $\Sigma^i_a = \Sigma^j_b$.
7. bad_7 is the event that there exists construction query $(\mathsf{eval}, i, M^i_a, T^i_a)$ and $(\mathsf{eval}, j, M^j_b, T^j_b)$ such that their block cipher keys are the same, namely, $K_i = K_j$; and their right input of the block cipher are the same, namely, $\Psi^i_a = \Psi^j_b$.

$\neg\mathsf{bad}_8$ guarantees that if there exists user i whose block cipher key is the same as the hash key of a user j, then in their construction queries, the inputs of the block cipher for user i are distinct from the input of the hash function for user j.

8. bad_8 is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, j, M_b^j, T_b^j)$ such that the block cipher key of i is the same as the left hash key of j, and the left hash output Σ_a^i is a message block in M_b^j or a proceed block of M_b^j during the hashing process in H; or the block cipher key of i is the same as the right hash key of j, and the right hash output Ψ_a^i is a message block in M_b^j or a proceed block of M_b^j during the hashing process in H.

$\neg\mathsf{bad}_9$ guarantees that for each user i and each tuple of its construction query, at least one of inputs of the block cipher is fresh.

9. bad_9 is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ such that both their left hash output and right hash output are the same, namely, $\Sigma_a^i = \Sigma_b^i$ and $\Psi_a^i = \Psi_b^i$.

$\neg\mathsf{bad}_{10} \wedge \neg\mathsf{bad}_{11}$ guarantees that, for each user i, the partial function Φ (Algorithm 1) will always not abort and simulate block cipher output (U_a^i, Q_a^i) for each tuple of construction query in the ideal world.

10. bad_{10} is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ 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 $U_a^i = Q_b^i$.
11. bad_{11} is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ 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 $Q_a^i = Q_b^i$.

$\neg\mathsf{bad}_{12} \wedge \neg\mathsf{bad}_9$ guarantees that for each user i and each triple of its construction query, at least one of inputs of the block cipher is fresh.

12. bad_{12} is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$, $(\mathsf{eval}, i, M_b^i, T_b^i)$ and $(\mathsf{eval}, i, M_c^i, T_c^i)$ such that

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

$\neg\mathsf{bad}_{13} \wedge \neg\mathsf{bad}_{14} \wedge \neg\mathsf{bad}_{10} \wedge \neg\mathsf{bad}_{11}$ guarantees that, for each user i, the partial function Φ (Algorithm 1) will always not abort and simulate block cipher output (U_a^i, Q_a^i) for any construction query in the ideal world.

13. bad_{13} is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$, $(\mathsf{eval}, i, M_b^i, T_b^i)$ and $(\mathsf{eval}, i, M_c^i, T_c^i)$ such that

$$\left[\Sigma_a^i = \Sigma_b^i \text{ or } \Sigma_a^i = \Psi_b^i\right] \text{ and } \left[U_a^i = U_c^i \text{ or } U_a^i = Q_c^i\right].$$

14. bad_{14} is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$, $(\mathsf{eval}, i, M_b^i, T_b^i)$ and $(\mathsf{eval}, i, M_c^i, T_c^i)$ such that

$$\left[\Psi_a^i = \Sigma_b^i \text{ or } \Psi_a^i = \Psi_b^i\right] \text{ and } \left[Q_a^i = U_c^i \text{ or } Q_a^i = Q_c^i\right].$$

$\neg\mathsf{bad}_{15}$ guarantees that there are no more than n users sharing the same ideal cipher key.

15. bad_{15} is the event that there exists n users who share the same block cipher key.

Bounding Probability of Bad Transcript. We now compute the probability of a bad script occurred in the ideal world. To compute it, we first bound the probability of each bad event occurred.

1. For a fixed i, $\Pr\left[K_i = K^i_{h,1} \vee K_i = K^i_{h,2}\right] \le \frac{2}{2^k}$. By the union bound, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_1\right] \le \frac{2u}{2^k}.$$
2. bad_2, in other words, says there exists i and $d \in \{1, 2\}$ such that both $(K_i \in \{K_j, K^j_{h,1}, K^j_{h,2}\}$ for a j or $J = K_i$ for a primitive query $(\mathsf{prim}, J, x, y, *)$ or $J = K$ for a primitive query $(\mathsf{prim}, J, x, y, *))$ and $(K^j_{h,d} \in \{K_j, K^j_{h,1}, K^j_{h,2}\}$ for a j or $J = K_i$ for a primitive query $(\mathsf{prim}, J, x, y, *)$ or $J = K^j_{h,d}$ for a primitive query $(\mathsf{prim}, J, x, y, *))$. By the union bound, we simply have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_2\right] \le \frac{u(3u+p)(6u+2p)}{2^{2k}}.$$
3. bad_3, in other words, says there exists J, i and $d \in \{1, 2\}$ such that $J = K^i_{h,d}$ where J is the block cipher key used in a primitive query and $K^i_{h,d}$ is the hash key used by a user; and there exists a and a message block x such that $x \in M^i_a \bigcup \Sigma^i_a \bigcup \Psi^i_a$ and x is query by the primitive query $(\mathsf{prim}, J, x, y, *)$. Then, if $p + uq_m l \le 2^{n-2}$, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_3\right] \le \frac{2upq_m l}{2^{k+n}}.$$
4. bad_4, in other words, says there exists i, a, J, x in the distinguisher $\mathcal{D}$'s transcript such that $J = K_i$; and $x = \Sigma^a_i$ or $x = \Psi^a_i$. Then, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_4 | \neg\mathsf{bad}_2\right] \le \frac{2upq_m\delta_1}{2^k}.$$
5. bad_5, in other words, says there exists i, a, J, x in the distinguisher $\mathcal{D}$'s transcript such that $J = K_i$; and $y = U^i_a$ or $y = Q^i_a$. The event $U^i_a = y$ or $Q^i_a = y$ is the same as $\Phi_{K_i}(U^i_a) = y$ or $\Phi_{K_i}(Q^i_a) = y$ (Recall Φ is the partial function used to simulate a random permutation, see Algorithm 1). Then, if $p + uq_m l \le 2^{n-2}$, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_5 | \neg\mathsf{bad}_2\right] \le \frac{8upq_m}{2^{k+n}}.$$
6. We compute the probability of bad_6 occurs conditioned on the bad event bad_2 not happening. In other words, the condition makes sure the two construction query use different left hash key. Since the hash function is δ_1-regular, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_6 | \neg\mathsf{bad}_2\right] \le \frac{u^2 q_m^2 \delta_1}{2^k}.$$

7. We compute the probability of bad_7 occurs conditioned on the bad event bad_2 not happening. Following the same method of (6), we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_7 | \neg \mathsf{bad}_2\right] \leq \frac{u^2 q_m^2 \delta_1}{2^k}.$$

8. To compute the probability of bad_8 occurs, we first consider the first case and the second case follows the same calculation. Here we call the first case bad_{8a}, which says, there exists i, j, a, b in the distinguisher $\mathcal{D}$'s transcript such that $K_i = K_{h,1}^j$ and Σ_a^i is appeared in either M_b^j or the hashing process with key $K_{h,1}^j$. Then we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_8\right] = 2 \Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{8a}\right] \leq \frac{2u^2 q_m^2 l \delta_1}{2^k}.$$

9. Since we assume the distinguisher $\mathcal{D}$ will not issue the same construction query twice, and the hash function is δ_2-AU, we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_9\right] \leq u q_m^2 \delta_2^2.$$

10. We can bound the probability of bad_{10} as
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{10}\right] \leq \frac{2u q_m^2 (\delta_1 + \delta_2)}{2^n}.$$

11. Similar to event bad_{10}, we can bound the probability of bad_{11} as
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{11}\right] \leq \frac{2u q_m^2 (\delta_1 + \delta_2)}{2^n}.$$

12. Similar to event bad_9, we can bound the probability of bad_{12} as
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{12}\right] \leq 2u q_m^3 \delta_2^2.$$

13. Similar to event bad_{10}, we can bound the probability of bad_{13} as
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{13}\right] \leq \frac{2u q_m^3 (\delta_1 + \delta_2)}{2^n}.$$

14. Similar to event bad_{13}, we can bound the probability of bad_{14} as
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{14}\right] \leq \frac{2u q_m^3 (\delta_1 + \delta_2)}{2^n}.$$

15. bad_{15}, in other words, says, there exists $i_1, \ldots, i_n \in [u]$ such that $K_{i_1} = \cdots = K_{i_n}$. So we have
$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}_{15}\right] = \frac{\binom{u}{n}}{2^{k(n-1)}} \leq \frac{u^2}{2^{k+n}}.$$

By the union bound, the overall probability of the bad event is bounded as follows:

$$\begin{aligned}\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{bad}\right] \leq \frac{2u}{2^k} + \frac{u(3u+p)(6u+2p)}{2^{2k}} + \frac{2upq_m l}{2^{k+n}} + \frac{2upq_m \delta_1}{2^k} \\ + \frac{8upq_m}{2^{k+n}} + \frac{4u^2q_m^2 l\delta_1}{2^k} + 3uq_m^3\delta_2^2 + \frac{8uq_m^3(\delta_1+\delta_2)}{2^n} + \frac{u^2}{2^{k+n}}. \quad (34)\end{aligned}$$

Good Transcript Analysis. The following analysis is built upon the analysis presented in [42], highlighting the modifications relevant to the context of our paper. Let τ be a good transcript. Recall a transcript is good if it is not bad, namely $\tau \notin \mathsf{bad}$. Note that for any good transcript, for each construction query, at least one of hash output, namely Σ_a^i and Ψ_a^i, is fresh. So the number of distinct hash output value is $|\mathbb{Q}(J)| + |F(J)|$, and number of duplicated hash output value is $|\mathbb{Q}(J)| - |F(J)|$. We use g to denote the number of distinct hash output values that has at least one duplication. Recall the definition of $\mathbb{Q}(J), F(J)$ from Algorithm 1. In the ideal world, we have

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

In the real world, we have

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

Then, we have

$$\begin{aligned}
\frac{\Pr\left[\mathsf{T}_{\mathsf{re}} = \tau\right]}{\Pr\left[\mathsf{T}_{\mathsf{id}} = \tau\right]} \geq (N-1)^{uq_m} \prod_{J \in \{0,1\}^k} \frac{|S(J)|\,(N - 2\,|F(J)|)_g}{(N)_{|Q(J)|+|F(J)|+g}} \\
&\geq \prod_{J \in \{0,1\}^k} \frac{(N-1)^{|Q(J)|}(N - 2\,|F(J)|)_g}{(N)_{|Q(J)|+|F(J)|+g}} \cdot |S(J)| \\
&\geq \prod_{J \in \{0,1\}^k} \frac{(N-1)^{|Q(J)|}(N - 2\,|F(J)|)_g}{(N)_{|Q(J)|+|F(J)|+g}} \frac{(N)_{2|F(J)|}}{(N-1)^{|F(J)|}} \\
&\quad \times \left(1 - \frac{2\,|F(J)|^2}{N^2} - \frac{128\,|F(J)|^3}{N^3} - \frac{8(n+1)^3}{3N^2}\right) \quad \text{(by Theorem 5)} \\
&\geq \prod_{J \in \{0,1\}^k} \frac{(N-1)^{|Q(J)|-|F(J)|}}{(N - 2\,|F(J)| - g)_{|Q(J)|-|F(J)|}} \\
&\quad \times \left(1 - \frac{2\,|F(J)|^2}{N^2} - \frac{128\,|F(J)|^3}{N^3} - \frac{8(n+1)^3}{3N^2}\right) \\
&\geq \prod_{J \in \{0,1\}^k} \frac{(N-1)^{|Q(J)|-|F(J)|}}{(N - 2\,|F(J)| - g)_{|Q(J)|-|F(J)|}} \\
&\quad \times \left(1 - \frac{2n^2q_m^2}{N^2} - \frac{128n^3q_m^3}{N^3} - \frac{8(n+1)^3}{3N^2}\right) \\
&\qquad\qquad (\because |F(J)| \leq nq_m, \text{ guaranteed by } \neg\mathsf{bad}_{15}) \\
&\geq 1 - \frac{2un^2q_m^2}{N^2} - \frac{128un^3q_m^3}{N^3} - \frac{8(n+1)^3u}{3N^2}, \qquad (35)
\end{aligned}$$

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

Conclude the Proof. From Equations (34) and (35), define

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

and

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

Then by Lemma 20, we conclude that

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

E.4 Proof of Theorem 9

The general idea of the proof follows the proof of [24, 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 [24]. 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 modify the calculations associated with certain bad events presented in [24], and also lead to consequential shifts in the final statement (Theorem 9). 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 9 trivially holds. Hence in the following, we prove Theorem 9 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. 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. During the computation of $(\mathsf{eval}, i, M_a^i, T_a^i)$, let $\varSigma_a^i, \varPsi_a^i$ be the internal outputs of hash function H, namely $\varSigma_a^i = \mathsf{H}^1_{K_{h,1}}(M_a^i)$ and $\varPsi_a^i = \mathsf{H}^2_{K_{h,2}}(M_a^i)$, respectively. Let U_a^i, Q_a^i be the outputs of ideal cipher E deployed in DbHtS with inputs $\varSigma_a^i$ and $\varPsi_a^i$, namely $U_a^i = \mathsf{E}(K_i, 0\|\varSigma_a^i)$ and $Q_a^i = \mathsf{E}(K_i, 1\|\varPsi_a^i)$, 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.

Algorithm 2 Offline oracle in the ideal world

```
$(K^i_{h,1}, K^i_{h,2})_{i\in[u]} \leftarrow_\$ \mathcal{K}_h \times \mathcal{K}_h$
$(K_i)_{i\in[u]} \leftarrow_\$ \{0,1\}^k$
$(\Sigma^i_a, \Psi^i_a)_{(i,a)\in[u]\times[q_m]} \leftarrow (\mathsf{H}^1_{K_{h,1}}(M^i_a), \mathsf{H}^2_{K_{h,2}}(M^i_a))_{(i,a)\in[u]\times[q_m]}$
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
$\mathbb{Q}^= \overset{\text{def}}{=} \{(i,a) \in [u] \times [q_m] : \exists(\mathsf{prim}, K_i, x, y, *); \forall(\mathsf{prim}, K_i, x, y, *), x \neq \Sigma^i_a, x \neq \Psi^i_a\}$
$\mathbb{I}^= \overset{\text{def}}{=} \{i \in [u] : (i,*) \in \mathbb{Q}^=\} = \mathbb{I}^=_{i_1} \sqcup \cdots \sqcup \mathbb{I}^=_{i_r}$    ▷ $i \in \mathbb{I}^=_{i_j}$ 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 $\Sigma^i_a$ 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 $\mathsf{Dom}(K_{i_j}) \overset{\text{def}}{=} \{x : (\mathsf{prim}, K_{i_j}, x, y, *)\}$ and $\mathsf{Ran}(K_{i_j}) \overset{\text{def}}{=} \{y : (\mathsf{prim}, K_{i_j}, x, y, *)\}$
    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 $\mathsf{Dom}(K_{i_j}) \leftarrow \mathsf{Dom}(K_{i_j}) \bigcup \{0\|\Sigma^i_a, 1\|\Psi^i_a\}, \mathsf{Ran}(K_{i_j}) \leftarrow \mathsf{Ran}(K_{i_j}) \bigcup \{U^i_a, Q^i_a\}$
$\mathbb{Q}^{\neq} \overset{\text{def}}{=} \{(i,a) \in [u] \times [q_m] : \forall(\mathsf{prim}, J, x, y, *), J \neq K_i\}$
$\mathbb{I}^{\neq} \overset{\text{def}}{=} \{i \in [u] : (i,*) \in \mathbb{Q}^{\neq}\} = \mathbb{I}^{\neq}_{i_1} \sqcup \cdots \sqcup \mathbb{I}^=_{i_{r'}}$,    ▷ $i, j \in \mathbb{I}^{\neq}_{i_j}$ 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 $\mathcal{S}_{i_j} \overset{\text{def}}{=} \{\bigcup_{i\in\mathbb{I}^{\neq}_{i_j}, a\in[q_m]} \{Z^i_{a,1}, Z^i_{a,2}\} \in (\{0,1\}^n)^{|\widetilde{\Sigma^{i_j}}|+|\widetilde{\Psi^{i_j}}|} : Z^i_{a,1} \oplus Z^i_{a,2} = T^i_a\}$
    return $(\Sigma^i_a, \Psi^i_a, U^i_a, Q^i_a)_{(i,a)\in[u]\times[q_m]}$
```

In the real world, after the distinguisher $\mathcal{D}$ finishes all its queries, we will further give the following information to the distinguisher $\mathcal{D}$:

1. the keys (K^i_h, 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 construction query a.

In the ideal world, we will instead give the distinguisher $\mathcal{D}$:

1. $(K^i_h, K_i) \leftarrow_\$ \{0,1\}^{2k} \times \{0,1\}^k$, independent of its queries.
2. the dummy values $(\Sigma^i_a, \Psi^i_a, U^i_a, Q^i_a)$ computed by the simulation oracle given in Algorithm 2. Note that the simulation oracle is almost the same as [24, Figure 4.3]. The difference is that we remove Bad-Tag event since we are assuming the fine-tuned ideal world while there is an additional bad event, dubbed 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 $(\Sigma_a^i, \Psi_a^i, U_a^i, Q_a^i)$ for each of u users and each of their q_m construction queries,
3. and p primitive queries and uq_m construction queries.

Bad Transcript Analysis and Interpretations. We now define bad transcripts. Let T_{id} and T_{re} be random variables following the distribution of a transcript in the real world and the ideal world, respectively. We call the transcript *bad* if any bad events happen, and *good* if it is not bad.

1. BadK is the event that there exists distinct user i_1 and i_2 such that they share the same block-cipher key and at least one of left hash key and right hash key.
2. Bad1 is the event that there exists a construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and a primitive query $(\mathsf{prim}, J, x, y, *)$ such that the block cipher key K_i collides with the ideal cipher key J; and at least one of hash output of the construction query collide with x, which is either the input of the primitive query or output of the inverse primitive query.

Bad2 is the union of event B.21 and B.22, which are described below.

3. B.21 is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ from a same user i such that they have the same MAC tag and at least one of hash output is collided.
4. B.22 is the event that there exists construction query $(\mathsf{eval}, i_1, M_a^{i_1}, T_a^{i_1})$ and $(\mathsf{eval}, i_2, M_b^{i_2}, T_b^{i_2})$ from distinct users i_1, i_2 such that they have the same block cipher key and at least one of hash output is collided.

Bad3 is the union of event B.31 and B.32, which are described below.

5. Let $L_1 = \frac{q_m^2 \delta}{2} + \frac{q_m \delta^{\frac{1}{4}}}{2}$. B.31 is the event that there exists a user i such that it has L_1 construction query tuple, which is with the same left hash output value. In other words,
$$\left|\{(a,b) : a < b \wedge \Sigma_a^i = \Sigma_b^i\}\right| \geq L_1.$$
6. B.32 is the event that there exists a user i such that it has L_1 construction query tuple, which is with the same right hash output value. In other words,
$$\left|\{(a,b) : a < b \wedge \Psi_a^i = \Psi_b^i\}\right| \geq L_1.$$

Bad4 is the union of event B.41, B.42, and B.43, which are described below.

7. B.41 is the event that there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ such that both their left hash output and right hash output are the same, namely, $\Sigma_a^i = \Sigma_b^i$ and $\Psi_a^i = \Psi_b^i$.
8. B.42 is the event that there exists a user i and its four distinct construction queries indexed by a, b, c, d such that
$$\left[\Sigma_a^i = \Sigma_b^i\right] \text{ and } \left[\Psi_b^i = \Psi_c^i\right] \text{ and } \left[\Sigma_c^i = \Sigma_d^i\right].$$

Note that since our DbHtS construction employs the domain separation, the hash outputs of two hash functions never collide, e.g., Σ_a^i cannot collide with Ψ_b^i. [24] achieves the same effect by assuming the hash function is cross-collision resistant.

9. B.43 is the event that there exists a user i and its four distinct construction queries indexed by a, b, c, d such that

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

Bad5 is described below. Looking ahead, $\neg$Bad5 guarantees that when we construct a bipartite graph for each user's construction queries, the number of component with size larger than 2 is at most L_2. The details and context can be found in good transcript analysis.

10. Let $L_2 = q_m^{\frac{1}{3}} \delta^{\frac{1}{2}} 2^{\frac{2}{3}n}$. Bad5 is the event that there exists user i with at least L_2 construction query triple (a, b, c) such that $\Sigma_a^i = \Sigma_b^i$ and $\Psi_b^i = \Psi_c^i$. In other words,

$$\left|\{(a, b, c) \in [q_m]^{*3} : \Sigma_a^i = \Sigma_b^i \wedge \Psi_b^i = \Psi_c^i\}\right| \geq L_2.$$

Bad-Samp is the event that there exists a user i and a primitive query $(\mathsf{prim}, J, x, y, *)$ such that the ideal cipher key J collides with user i's block cipher key K_i; and the simulation oracle [line 12, Algorithm 2] fails to simulate block cipher output Q_a^i for user i's construction query. We bound Bad-Samp by the union of events BS1 and BS2 described below.

11. BS1 is the event that there exists a primitive query $(\mathsf{prim}, J, x, y, *)$ and a user i such that the ideal cipher key J collides with user i's block cipher key K_i; and y collides with the simulation oracle [line 12, Algorithm 2]'s output Q_a^i.
12. BS2 is the event that there exists a primitive query $(\mathsf{prim}, J, x, y, *)$ and two users i, i' such that the ideal cipher key J collides with the block cipher key of both user i and i'; and simulated block cipher output Q_a^i collides with $Q_b^{i'}$, where both Q_a^i and $Q_b^{i'}$ are both simulated by the simulation oracle [line 12, Algorithm 2]. Note that i and i' are not necessarily distinct.

Bounding Probability of Bad Transcript. We now compute the probability of a bad script occurred in the ideal world. To compute it, we first bound the probability of each bad event occurred.

1. BadK, in other words, says there exists distinct $i_1, i_2 \in [u]$ such that $K_{i_1} = K_{i_2} \wedge (K_{h,1}^{i_1} = K_{h,1}^{i_2} \vee K_{h,2}^{i_1} = K_{h,2}^{i_2})$. Since both block cipher key and hash key are independently sampled, we have

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

2. $\mathsf{Bad1}$, in other words, says there exists a construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and a primitive query $(\mathsf{prim}, J, x, y, *)$ such that

   $$[K_i = J] \text{ and } \left[\varSigma_a^i = x \text{ or } \varPsi_a^i = x\right].$$

   Since the hash function is δ-regular, we have

   $$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{Bad1}\right] \leq \frac{2upq_m\delta}{2^k}.$$

3. $\mathsf{B.21}$, in other words, says there exists construction query $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ such that

   $$\left[T_a^i = T_b^i\right] \text{ and } \left[\varSigma_a^i = \varSigma_b^i \text{ or } \varPsi_a^i = \varPsi_b^i\right].$$

   We have

   $$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.21}\right] \leq \frac{2uq_m^2\delta}{2^n}.$$

4. $\mathsf{B.22}$, in other words, says there exists construction query $(\mathsf{eval}, i_1, M_a^{i_1}, T_a^{i_1})$ and $(\mathsf{eval}, i_2, M_b^{i_2}, T_b^{i_2})$ such that

   $$[K_{i_1} = K_{i_2}] \text{ and } \left[\varSigma_a^{i_1} = \varSigma_b^{i_2} \text{ or } \varPsi_a^{i_1} = \varPsi_b^{i_2}\right].$$

   We have

   $$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.22}\right] \leq \frac{2uq_m^2\delta}{2^k}.$$

5. We then bound the probability of event $\mathsf{Bad2}$ from $\mathsf{B.21}$ and $\mathsf{B.22}$ using the union bound. We have

   $$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{Bad2}\right] \leq \frac{2uq_m^2\delta}{2^n} + \frac{2uq_m^2\delta}{2^k}.$$

6. We now upper bound the probability of event $\mathsf{B.31}$. Let $\mathbb{I}_{a,b}^i$ be the indicator random variable which takes the value 1 if $\varSigma_a^i = \varSigma_b^i$; 0 otherwise. In other words, given two construction queries $(\mathsf{eval}, i, M_a^i, T_a^i)$ and $(\mathsf{eval}, i, M_b^i, T_b^i)$ for the same user i, $\mathbb{I}_{a,b}^i$ is 1 if two queries have the same left hash output, otherwise 0. Let $\mathbb{I}^i = \sum_{a<b} \mathbb{I}_{a,b}^i$, which counts the number of construction queries tuple having the same left hash output. Since the hash function is δ-regular, we have

   $$\mathbf{Ex}\left[\mathbb{I}^i\right] = \sum_{a<b} \Pr\left[\varSigma_a^i = \varSigma_b^i\right] \leq \frac{q_m^2\delta}{2}.$$

   Also, since the hash function is δ-AU$^{(2)}$, we have

   $$\mathbf{Var}\left[\mathbb{I}^i\right] \leq \mathbf{Ex}\left[(\mathbb{I}^i)^2\right] = \mathbf{Ex}\left[\left(\sum_{a<b} \mathbb{I}_{a,b}^i\right)\left(\sum_{c<d} \mathbb{I}_{c,d}^i\right)\right] \leq \frac{q_m^4\delta^2}{4}.$$

Using the union bound on all event $\mathbb{I}^i \geq L_1$ for each user i, we have

$$\begin{aligned}
\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.31}\right] &\leq \sum_{i\in[u]} \Pr\left[\mathbb{I}^i \geq L_1\right] \\
&\leq \frac{uq_m^4\delta^2}{(2L_1 - q_m^2\delta)^2} && \text{(by Lemma 3)} \\
&= \frac{uq_m^4\delta^2}{q_m^2\delta^{\frac{1}{2}}} && (\because \text{Plug in } L_1) \\
&\leq uq_m^2\delta^{\frac{3}{2}}.
\end{aligned}$$

7. Similar to event B.31, we can bound the probability of event B.32 as

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.32}\right] \leq uq_m^2\delta^{\frac{3}{2}}.$$

8. We then bound the probability of event Bad3 from B.31 and B.32 using the union bound. We have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.32}\right] \leq 2uq_m^2\delta^{\frac{3}{2}}.$$

9. Since we assume the distinguisher $\mathcal{D}$ will not issue the same construction query twice, and the hash function is δ-AU, we have

$$\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.41}\right] \leq \frac{uq_m^2\delta^2}{2}.$$

10. $\neg$B.31, namely, the number of construction query tuple with the same left hash output is at most L_1, guarantees that the number of quadruples (a, b, c, d) such that $\left[\varSigma_a^i = \varSigma_b^i\right]$ and $\left[\varSigma_c^i = \varSigma_d^i\right]$ is at most L_1^2. Since the hash function is δ-AU, we have

$$\begin{aligned}
\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.42} | \neg\mathsf{B.31}\right] &\leq \sum_{i\in[u]} L_1^2\delta \\
&= \left(\frac{q_m^2\delta}{2} + \frac{q_m\delta^{\frac{1}{4}}}{2}\right)^2 \delta && (\because \text{Plug in } L_1) \\
&\leq \frac{1}{2}q_m^4\delta^3 + \frac{1}{2}q_m^2\delta^{\frac{3}{2}} && \text{(by Lemma 5)} \\
&\leq q_m^2\delta^{\frac{3}{2}}. && (\because q_m^2\delta^{\frac{3}{2}} < 1)
\end{aligned}$$

Further,

$$\begin{aligned}
\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.42}\right] \leq\ & \Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.42} | \neg\mathsf{B.31}\right] + \Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.31}\right] \\
\leq\ & 2q_m^2\delta^{\frac{3}{2}}.
\end{aligned}$$

Note that the summation of $\Pr\left[\mathsf{T}_{\mathsf{id}} \in \mathsf{B.31}\right]$ 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.

11. Similar to event $\mathsf{B.42}$, we can bound the probability of event $\mathsf{B.43}$ as

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{B.43}\right] \leq 2q_m^2\delta^{\frac{3}{2}}.$$

12. We then bound the probability of event $\mathsf{Bad4}$ from $\mathsf{B.41}$, $\mathsf{B.42}$, and $\mathsf{B.43}$ using the union bound. We have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{Bad4}\right] \leq \frac{uq_m^2\delta^2}{2} + 4uq_m^2\delta^{\frac{3}{2}} \leq \frac{9uq_m^2\delta^{\frac{3}{2}}}{2}.$$

13. We now bound the probability of event $\mathsf{Bad5}$. Let $\mathbb{I}^i_{a,b,c}$ be the indicator random variable which takes the value 1 if $\Sigma^i_a = \Sigma^i_b \wedge \Psi^i_b = \Psi^i_c$; 0 otherwise. Let

$$\mathbb{I}^i = \sum_{(a,b,c)\in[q_m]^{*3}} \mathbb{I}^i_{a,b,c}.$$

Since the hash function is δ-AU$^{(2)}$, we have

$$\mathbf{Ex}\left[\mathbb{I}^i\right] = \sum_{(a,b,c)\in[q_m]^{*3}} \Pr\left[\Sigma^i_a = \Sigma^i_b \wedge \Psi^i_b = \Psi^i_c\right] \leq q_m^3\delta^2,$$

Then, by Lemma 2, we have

$$\begin{aligned}\Pr\left[\mathsf{T_{id}} \in \mathsf{Bad5}\right] &\leq \sum_{i\in[u]} \Pr\left[\mathbb{I}^i \geq L_2\right] \\ &\leq \frac{uq_m^3\delta^2}{L_2} \\ &= \frac{uq_m^{\frac{8}{3}}\delta^{\frac{3}{2}}}{2^{\frac{2}{3}n}}. && (\because \text{Plug in } L_2)\end{aligned}$$

14. We can bound the probability of $\mathsf{BS1}$ as

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS1}\right] \leq \frac{2upq_m}{2^{n+k}}.$$

15. We can bound the probability of $\mathsf{BS2}$ as

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \leq \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

To prove the inequality, we show for every setting of u, q_m the inequality holds. We introduce a new variable $q = uq_m$ to partition the case of u, q_m. We first note the following inequality holds

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \leq \sum_{i=1}^{u} \frac{pq_m^2}{2^{n+k}} = \frac{upq_m^2}{2^{n+k}} = \frac{pqq_m}{2^{n+k}}.$$

If $q \leq 2^{n/2}$, then we have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \leq \frac{pqq_m}{2^{n+k}} \leq \frac{pq_m}{2^{\frac{n}{2}+k}} \leq \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

On the other hand, if $q > 2^{n/2}$, we start with considering the first $\frac{q}{2^{\frac{n}{2}}}$ users. Similarly to [24, 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

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{Aux}\right] \leq \frac{\left(\frac{q}{2^{\frac{n}{2}}}\right) p}{2^k} \leq \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

If $u \leq \frac{q}{2^{\frac{n}{2}}}$, $\mathsf{BS2} \subseteq \mathsf{Aux}$, so we have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \leq \Pr\left[\mathsf{T_{id}} \in \mathsf{Aux}\right] \leq \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

Otherwise, $q = uq_m \geq \frac{q}{2^{\frac{n}{2}}} q_m$, which says $q_m \leq 2^{n/2}$. Then we have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \leq \frac{pqq_m}{2^{n+k}} \leq \frac{pq}{2^{\frac{n}{2}+k}} = \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

Since for all cases of u, q_m, the inequality holds. We conclude that

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{BS2}\right] \leq \frac{upq_m}{2^{\frac{n}{2}+k}}.$$

16. We then bound the probability of event $\mathsf{Bad\text{-}Samp}$ from $\mathsf{BS1}$ and $\mathsf{BS2}$ using the union bound. We have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{Bad\text{-}Samp}\right] \leq \frac{2upq_m}{2^{n+k}} + \frac{upq_m}{2^{\frac{n}{2}+k}} \leq \frac{3upq_m}{2^{\frac{n}{2}+k}}.$$

Define $\mathsf{bad} = \mathsf{BadK} \vee \mathsf{Bad1} \vee \mathsf{Bad2} \vee \mathsf{Bad3} \vee \mathsf{Bad4} \vee \mathsf{Bad5} \vee \mathsf{Bad\text{-}Samp}$, and summing up the probability of all bad events, we have

$$\Pr\left[\mathsf{T_{id}} \in \mathsf{bad}\right] \leq \frac{2u^2}{2^{2k}} + \frac{2upq_m\delta}{2^k} + \frac{2u^2q_m^2\delta}{2^k} + 8uq_m^2\delta^{\frac{3}{2}} + \frac{3upq_m}{2^{\frac{n}{2}+k}}. \qquad (36)$$

Good Transcript Analysis. The following analysis computes a lower bound of the ratio $\frac{\Pr[\mathsf{T_{re}}=\tau]}{\Pr[\mathsf{T_{id}}=\tau]}$, where τ is a good transcript. Recall $\mathsf{T_{id}}$ and $\mathsf{T_{re}}$ are random variables following the distribution of a transcript in the real world and the ideal world, respectively. We call the transcript *bad* if any bad events happen, and *good* if it is not bad.

We first consider transcripts for construction queries indexed by $\mathbb{Q}^=$. Recall $\mathbb{Q}^=$ is a set of construction queries for users whose block cipher key collide with one or more ideal cipher keys used in primitive queries. Formally,

$$\mathbb{Q}^= \overset{\text{def}}{=} \left\{(i,a) \in [u] \times [q_m] : \exists(\mathsf{prim}, K_i, x, y, *); \forall(\mathsf{prim}, K_i, x, y, *), x \neq \varSigma_a^i, x \neq \varPsi_a^i\right\}$$

as defined in Algorithm 2. Recall r is the total number of distinct block cipher keys that collide with ideal cipher keys, and we use $i_1, \cdots, i_r$ to represent the r class of collided block cipher keys. Further, $\mathbb{I}_{i_1}^=, \cdots, \mathbb{I}_{i_r}^=$ are the corresponding set of users using the collided block cipher keys. For each $j \in [r]$ and each $i \in \mathbb{I}^{=}_{i_j}$, we consider the internal value sequence

$$(U^i_1, \dots, U^i_{q_m}), (Q^i_1, \dots, Q^i_{q_m}).$$

From this sequence, we construct a bipartite graph G_i, where the nodes in one partition represent values U^i_a and the nodes in the other represent Q^i_a. We connect the node representing U^i_a and Q^i_a with an edge labeled with T^i_a, where $U^i_a \oplus Q^i_a = T^i_a$. If $U^i_a = U^i_b$ where $a \neq b$, then we merge the corresponding nodes into a single one. We do the same thing if $Q^i_a = Q^i_b$ where $a \neq b$.

Since the transcript τ 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 $L_1 = \frac{q_m^2 \delta}{2} + \frac{q_m \delta^{\frac{1}{4}}}{2}$. Guaranteed by $\neg\mathsf{Bad5}$, the number of component with size larger than 2 is at most L_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 transcripts for construction queries indexed by $\mathbb{Q}^{\neq}$. Recall $\mathbb{Q}^{\neq}$ is a set of construction queries for users whose block cipher key doesn't collide with any ideal cipher keys used in primitive queries. r' is the total number of distinct block cipher keys that doesn't collide with any ideal cipher keys. Recall $\mathbb{I}^{\neq} \overset{\text{def}}{=} \{i \in [u] : (i, *) \in \mathbb{Q}^{\neq}\} = \mathbb{I}^{\neq}_{i_1} \sqcup \cdots \sqcup \mathbb{I}^{=}_{i'_r}$, as defined in Algorithm 2, is the union of set of users using r' class of block cipher keys. For each $j \in [r']$ and each $i \in \mathbb{I}^{\neq}_{i_j}$, we consider the internal value sequence

$$(U^i_1, \dots, U^i_{q_m}), (Q^i_1, \dots, Q^i_{q_m}).$$

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.

Recall we use s to denote the number of distinct ideal cipher keys and p_j to denote the number of primitive queries using the j-th ideal-cipher key, where $j \in [s]$. In the ideal world, we have

$$\Pr\left[\mathsf{T}_{\mathsf{re}} = \tau\right]$$

$$= \prod_{i=1}^{u} \frac{1}{2^{3k}} \cdot \left(\prod_{j=1}^{r} \frac{1}{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}^{=}_{i_j}} v_i\right)}} \right) \prod_{j \in [s] \setminus \{i_1, \dots, i_r\}} \frac{1}{(2^n)_{p_j}} \left(\prod_{j=1}^{r'} \frac{1}{(2^n)_{\left(\sum_{i \in \mathbb{I}^{\neq}_{i_j}} v'_i\right)}} \right).$$

In the real world, we have

$$\begin{aligned}
&\Pr\left[\mathsf{T}_{\mathsf{id}} = \tau\right] \\
&= \frac{1}{2^{nuq_m}} \prod_{i=1}^{u} \frac{1}{2^{3k}} \cdot \left(\prod_{j=1}^{r} \frac{1}{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}^{=}_{i_j}} w_i\right)}} \right) \prod_{j \in [s] \setminus \{i_1, \ldots, i_r\}} \frac{1}{(2^n)_{p_j}} \left(\prod_{j=1}^{r'} \frac{1}{\left|\mathcal{S}_{i_j}\right|} \right),
\end{aligned}$$

where $\mathcal{S}_{i_j}$ is defined in [line 17, Algorithm 2].

Plugging in the above two expressions, we have

$$\begin{aligned}
\frac{\Pr\left[\mathsf{T}_{\mathsf{re}} = \tau\right]}{\Pr\left[\mathsf{T}_{\mathsf{id}} = \tau\right]} &= 2^{nuq_m} \left(\prod_{j=1}^{r} \frac{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}^{=}_{i_j}} w_i\right)}}{(2^n)_{\left(p_j + \sum_{i \in \mathbb{I}^{=}_{i_j}} v_i\right)}} \right) \cdot \left(\prod_{j=1}^{r'} \frac{\left|\mathcal{S}_{i_j}\right|}{(2^n)_{\left(\sum_{i \in \mathbb{I}^{\neq}_{i_j}} v'_i\right)}} \right) \\
&\geq 2^{nuq_m} \left(\prod_{j=1}^{r} \frac{1}{\left(2^n - p_j - \sum_{i \in \mathbb{I}^{=}_{i_j}} w_i\right)_{\left(\sum_{i \in \mathbb{I}^{=}_{i_j}} (v_i - w_i)\right)}} \right) \cdot \left(\prod_{j=1}^{r'} \frac{\left(1 - \delta_{i_j}\right) \cdot (2^n)_{\left(\sum_{i \in \mathbb{I}^{\neq}_{i_j}} v'_i\right)}}{2^{\left(n \sum_{i \in \mathbb{I}^{\neq}_{i_j}} (v'_i - w'_i)\right)} \cdot (2^n)_{\left(\sum_{i \in \mathbb{I}^{\neq}_{i_j}} v'_i\right)}} \right),
\end{aligned}$$

(∵ Plug in Theorem 4)

where the last inequality is by plugging the lower bound of $\mathcal{S}_{i_j}$ from the mirror theory given in Theorem 4. The introduced new variable δ_{i_j} is defined as

$$\begin{aligned}
\delta_{i_j} &\overset{\text{def}}{=} \sum_{i \in \mathbb{I}^{\neq}_{i_j}} \frac{9q_{c,i}^2 \sum_{1 \leq k \leq \alpha_i} c_k^2}{8 \cdot 2^{2n}} + \frac{31 q_{c,i} q_m^2}{2^{2n}} + \frac{16 q_m^4}{2^{3n}} \\
&\leq \frac{9q_{c,i}^2 L_2^2}{8 \cdot 2^{2n}} + \frac{31 q_{c,i} q_m^2}{2^{2n}} + \frac{16 q_m^4}{2^{3n}}. && (\because \neg\mathsf{Bad5})
\end{aligned}$$

We here give more explanation of the expression of δ_{i_j}. Recall $\mathbb{I}^{\neq}_{i_j}$ is the set of users who share the same block cipher key and this key doesn't collide with any ideal cipher keys used in the primitive query. $q_{c,i}$ denotes the total number of edges in H_i's components with size larger than 2. α_i denotes the total number of components in H_i with size larger than 2 and c_k denotes the size of k-th component in the graph H_i.

Continuing on the above inequality, we further have

$$
\begin{aligned}
\frac{\Pr\left[\mathsf{T}_{\mathsf{re}} = \tau\right]}{\Pr\left[\mathsf{T}_{\mathsf{id}} = \tau\right]} &\geq \left(\prod_{j=1}^{r} \frac{2^{nq_m |\mathbb{I}^{=}_{i_j}|}}{\left(2^n - p_j - \sum_{i \in \mathbb{I}^{=}_{i_j}} w_i\right)_{\left(\sum_{i \in \mathbb{I}^{=}_{i_j}} (v_i - w_i)\right)}} \right) \cdot \left(\prod_{j=1}^{r'} \frac{2^{nq_m |\mathbb{I}^{\neq}_{i_j}|} \cdot \left(1 - \delta_{i_j}\right)}{2^{\left(n \sum_{i \in \mathbb{I}^{\neq}_{i_j}} (v'_i - w'_i)\right)}} \right) \\
&\geq 1 - \sum_{j=1}^{r'} \delta_{i_j} \\
&\geq 1 - \sum_{j=1}^{r'} \sum_{i \in \mathbb{I}^{\neq}_{i_j}} \left(\frac{9q_{c,i}^2 L_2^2}{8 \cdot 2^{2n}} + \frac{31 q_{c,i} q_m^2}{2^{2n}} + \frac{16 q_m^4}{2^{3n}} \right) \\
&\geq 1 - \sum_{j=1}^{r'} \sum_{i \in \mathbb{I}^{\neq}_{i_j}} \left(\frac{9 q_m^{\frac{8}{3}} \delta^{\frac{3}{2}}}{8 \cdot 2^{\frac{2}{3}n}} + \frac{31 q_m^4 \delta + 31 q_m^3 \delta^{\frac{1}{4}}}{2 \cdot 2^{2n}} + \frac{16 q_m^4}{2^{3n}} \right) \\
&\qquad\qquad (\because \neg\mathsf{Bad3} \wedge \neg\mathsf{Bad5}) \\
&\geq 1 - \left(\frac{9u q_m^{\frac{8}{3}} \delta^{\frac{3}{2}}}{8 \cdot 2^{\frac{2}{3}n}} + \frac{47 u q_m^3 \delta^{\frac{1}{4}}}{2^{2n}} \right) \qquad (37)
\end{aligned}
$$

Conclude the Proof. From Equations (36) and (37), define

$$
\epsilon_1 \overset{\text{def}}{=} \frac{9u q_m^{\frac{8}{3}} \delta^{\frac{3}{2}}}{8 \cdot 2^{\frac{2}{3}n}} + \frac{47 u q_m^3 \delta^{\frac{1}{4}}}{2^{2n}}
$$

and

$$
\epsilon_2 \overset{\text{def}}{=} \frac{2u^2}{2^{2k}} + \frac{2upq_m\delta}{2^k} + \frac{2u^2 q_m^2 \delta}{2^k} + 8u q_m^2 \delta^{\frac{3}{2}} + \frac{3upq_m}{2^{\frac{n}{2}+k}}.
$$

Then by Lemma 20, we conclude that

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