A Modular Approach to the Security Analysis of Two-Permutation Constructions

Yu Long Chen

imec-COSIC, KU Leuven, Belgium
 yulong.chen@kuleuven.be

Abstract. Constructions based on two public permutation calls are very common in today's cryptographic community. However, each time a new construction is introduced, a dedicated proof must be carried out to study the security of the construction. In this work, we propose a new tool to analyze the security of these constructions in a modular way. This tool is built on the idea of the classical mirror theory for block cipher based constructions, such that it can be used for security proofs in the ideal permutation model. We present different variants of this public permutation mirror theory such that it is suitable for different security notions.

We also present a framework to use the new techniques, which provides the bad events that need to be excluded in order to apply the public permutation mirror theory. Furthermore, we showcase the new technique on three examples: the Tweakable Even-Mansour cipher by Cogliati et al. (CRYPTO '15), the two permutation variant of the pEDM PRF by Dutta et al. (ToSC '21(2)), and the two permutation variant of the nEHtM_p MAC algorithm by Dutta and Nandi (AFRICACRYPT '20). With this new tool we prove the *multi-user* security of these constructions in a considerably simplified way.

Keywords: mirror theory, two permutation calls constructions, multiuser security, modular framework

1 Introduction

PERMUTATION-BASED CRYPTO. Following the selection of Keccak as the winner of the SHA-3 competition [2], cryptographic schemes based on public permutations gained a lot of traction in the research community. Nowadays, permutationbased constructions have become a trend in cryptography, and form a successful and full-fledged alternative to block-cipher based designs. Recently, in the first round of the ongoing NIST lightweight competition [1], 24 out of 57 submissions are based on public permutations, and 16 out of 24 permutation-based designs have been selected for the second round. These statistics show without a doubt the wide acceptance of permutation based designs in the community. The long line of research on the design of secret key constructions using public permutations originates with Even and Mansour [23], who designed a secret random permutation using a public permutation by xoring random keys to the input and output of this permutation. Later, their work was generalized to the Iterated Even-Mansour construction or the Key Alternating Cipher by [6,9,17,26], which is the backbone of today's block ciphers.

CONSTRUCTIONS BASED ON TWO PERMUTATION CALLS. In recent years, beyond birthday bound security has become a very popular topic in the field of symmetric key cryptography due to the rise of lightweight primitives. Admittedly, the state size of a permutation is typically very large: for example the SHA-3 permutation is of size 1600 bits, and a simple birthday bound secure construction built on SHA-3 would be secure up to an attack complexity of 2^{800} . However, this example permutation is on the extreme end: a big drawback of these big permutations is that they were not designed with lightweight applications in mind, the state of lightweight permutations such as SPONGENT [5] and PHOTON [25] can be as small as 88 and 100 bits, respectively. Hence, birthday bound secure constructions using these types of permutations are inadequate.

Due to the above-mentioned reason, beyond birthday-bound constructions based on two permutations are interesting to investigate. Indeed, it is possible to break any single-permutation construction by finding a collision between the input of the underlying permutation of the given construction and an input to the oracle of the public primitive, which happens with probability $\Omega(qp/2^n)$, where q is the number of queries to the construction and p is the number of queries to the underlying permutation. Constructions using more than two permutation calls can achieve even better security, however these are less efficient and difficult to analyze. On the other hand, constructions based on two permutation calls can achieve a resulting security bound of the form $O(qp^2/2^{2n})$, which is usually sufficient for most practical applications. In the last years, several types of constructions based on two permutation calls were proposed and analyzed. The most notable examples are the 2-round Even-Mansour cipher by Bogdanov et al. [6] and Chen et al. [8], the tweakable block cipher TEM by Cogliati et al. [15] and Dutta [19], the pseudorandom function SoEM by Chen et al. [11] and pEDM by Dutta et al. [22], the FPTP hash function by Chen and Tessaro [12], and the nonce-based MAC algorithm nEHtM_n by Dutta and Nandi [20]. Due to the similarity in the structures, the security proofs of these construction all share some relevance.

SINGLE VS MULTI-USER SECURITY. The security of most of the above constructions has been proven in the single-user setting. In practice, however, commonly used cryptographic constructions are usually deployed in contexts with a large number of users. An obvious question is to what extent the number of users will affect the security bound of these permutation-based constructions, or more specifically, can these constructions still have a security bound of the form $O(qp^2/2^{2n})$ in the multi-user setting? The concept of multi-user security was first introduced by Bellare, Boldyreva and Micali [3] in the context of public key cryptography, and was later extended by Biham [4] to symmetric key cryptanalysis. In the multi-user setting, attackers can adaptively distribute their q construction queries across multiple users with independent keys, and the attackers succeed as long as they can compromise one user key. Unfortunately, research on provable multi-user security for permutation-based constructions has been missing until now. The notable exceptions are the work of key-alternating ciphers by Mouha and Luykx for a single round [31], and Hoang and Tessaro for multiple rounds [26]. These works show that evaluating how security degrades as the number of users grows is a challenging technical problem, even when the security is known in the single-user setting. The generic reduction [7], however, does not help the constructions to maintain beyond birthday-bound security in the multi-user setting. For example, suppose the number of users is u, then simply applying the generic reduction to obtain multi-user security from single-user security introduces an extra factor u in the security bound. If the attacker only asks one query per user, then the security bound becomes

$$\frac{uqp^2}{2^{2n}} \le \frac{q^2p^2}{2^{2n}} \,,$$

which is only comparable to the $O(qp/2^n)$ security of one-call constructions.

Therefore, it appears that a dedicated analysis of the multi-user security is needed. Most security proofs in symmetric key cryptography today are based on the H-coefficients technique [9,33]. The idea behind this technique is that only a smaller number of transcripts are significantly more likely to appear in the ideal world than in the real world, namely: the bad transcripts. Usually, such proofs are performed as follows: (1) we first define a set of bad transcripts, (2) then the probability of observing bad transcripts in the ideal world is upper bounded, (3) and finally the ratio of observing good transcripts in the real and the ideal world is lower bounded. Note that points (1) and (3) are completely different problems than point (2). Since upper bounding the probability of the bad transcripts is a purely combinatorial problem and has little to do with cryptography, it relies heavily on the randomness of the generated keys. Defining the bad transcripts and lower bounding the ratio of the good transcripts depend, however, strongly on the way a particular construction is built. Unlike the case of block cipherbased constructions, where single-user security is usually proven in the standard model and multi-user security in the ideal-cipher model. For permutation-based constructions, ideal-permutation model analysis is used for both the single and multi-user settings. This raises the question whether or not it is possible to derive a modular approach that can be applied to constructions based on two permutation calls, which generically find the set of bad transcripts and a tight lower bound for the ratio of the corresponding good transcripts in both the single and *multi*-user settings, avoiding the long and involved dedicated analysis.

PATARIN'S MIRROR THEORY. Before we give an answer to this question, we recall Patarin's mirror theory [34], which is a very powerful but currently still unverified technique. Mirror theory is concerned with systems of $q_m \ge 1$ equations with $r \ge q_m$ unknowns of the form $v \oplus y = \lambda$, where v and y are two unknowns, and λ is a known value. The goal is to determine a lower bound on the number of possible solutions to the unknowns such that the solution does not contain collisions. Originally, Patarin derived mirror theory in order to prove the optimal *n*-bits security of the Xor of two secret Permutations construction (XoP). After the modernization by Mennink and Neves [30], who used mirror theory to prove the pseudorandom function security of EDM, EDMD and EWCDM, the applications of mirror theory seem to be increasing. For example, mirror theory was used to prove the security of the 2-round CLRW tweakable block cipher by Jha and Nandi [27], and to prove the security of the nonce-less MAC algorithms PolyMAC, SUM-ECBC, PMAC-Plus, 3kf9, and LightMAC-Plus by Kim et al. [28]. Datta et al. [18] were first to extend the mirror theory by including a system of $q_a \geq 1$ non-equations of the form $v \oplus y \neq \lambda$, and used it to prove the security of the nonce-based MAC algorithm DWCDM. Later, Dutta et al. [21], and Kim et al. [14] used it to prove the security of the nonce-based MAC algorithm nHEtM.

THE NEW IDEA. The reason why we can apply mirror theory to the above mentioned block cipher based constructions is because all these constructions can be viewed as the xor of two secret permutations. Note that when the permutations become public, the constructions have a structure that follows the Sum of Even-Mansour (SoEM) construction of Chen et al. [11]. Since the proofs of public permutation based constructions are all performed in the ideal permutation model, the attacker also gets access to the underlying permutations. Hence it is necessary to have an xor before the input and after the output of each of the permutation evaluations. Due to this important observation, almost all constructions based on two permutation calls can be viewed as the xor of two public permutations in the middle. This observation leads to the answer to the previous question, and the goal of this paper is to use the idea of mirror theory to build a modular technique that can be applied to all of the above mentioned permutation-based constructions.

1.1 Our Contribution

The goal of this paper is to derive a generic tool that can be used for the security analysis of constructions based on two public permutation calls and for different security notions. In order to do that, there are a few difficulties that we need to resolve. First of all, the traditional mirror theory is only suitable for the prf security, and we cannot simply apply it to the other settings. The second problem is that mirror theory does not consider primitive queries, and we need to include these queries in order to apply the theory for ideal permutation model proofs.

We solve the first problem using the approach proposed by Jha and Nandi [27] (see SUP Material A-B), and we derive different versions of public permutation mirror theory that are suitable for almost all popular security notions in symmetric key cryptography. On the other hand, since each primitive query defines exactly the input and the output of one permutation evaluation, we can solve the second problem by including a set of uni-variant affine equations of the form $v = \lambda$ and $y = \lambda$. Each uni-variate affine equation defines exactly one primitive query (where the input and output values of these queries are well defined).

With all these in mind, we derive two new theorems for the ideal public permutation model proofs. In Section 3.1, we explain the general setting for

traditional mirror theory, and our new technique for the ideal permutation model is defined and given in Section 3.2. We provide four theorems for different type of settings. Theorem 1 (a) is suitable for security notions such as sprp and tsprp, Theorem 1 (b) is suitable for security notions such as prf, weak prf, and (t)ccr [24], and Theorem 2 (b) is suitable for notions such as mac security (prf with non-equations). Since these three theorems already cover all currently know security notions, Theorem 2 (a) (sprp or tsprp with non-equations) has been added for the sake of completeness, as we are not aware of any notions on which it can be applied and will leave this for future research. We want to emphasize that our aim here is not to fix the proofs in the traditional mirror theory, but rather to focus on a new modular technique for permutation-based constructions. Hence, our technique focuses only on 2n/3-bit security, since this is a tight security bound (as we will see from the three examples) for constructions based on two permutations due to the presence of the primitive queries.

Another important contribution in this work is to provide a general framework to use the new techniques for the (multi-user) security analysis of constructions based on two independent permutation calls. The framework is given in Section 4. In general, to prove the security of constructions based on two permutation calls, one should first turn the query transcript into a system of bi-variate (non-)equations and uni-variate equations. This system of (non-)equations can be used to define the transcript graph (see Section 4.2). In our case, we decompose our graph into four subgraphs - the union of the components containing one "colliding vertex" (defined by an uni-variate affine equation), the union of "star" components, the set of isolated edges, and the set of isolated vertices. However, the graph may contain other types of components that prevent us from using the new technique, these components need to be excluded in our analysis. A very important part about this framework is that it provides a set of bad events that need to be considered in the security analysis to exclude these components. It seems, especially when non-equations also need to be considered, the analysis becomes very complex, which increases the chance to miss some bad events (see Section 4.3). The probability of these bad events must be upper bounded based on the randomness of the generated keys, sometimes difficult combinatorial techniques are required in the case of limited randomness. After these bad events are excluded, our new theory can be applied to the given system to determine a lower bound on the number of possible solutions to the unknowns, which in its turn defines the ratio of observing good transcripts in the real and the ideal world (see Section 4.4). This framework is useful for future designs, such that the future analysis will not miss any necessary bad events.

APPLICATIONS. We illustrate the new techniques by applying them to prove the *multi-user* security of Tweakable Even Mansour (TEM), permutation-based Encrypted Davies-Mayer (pEDM), and permutation-based version of nonce-based Enhance Hash-then-Mask (nEHtM_p). These three constructions are chosen because they use three important security notions in symmetric-key cryptography, namely tsprp, prf and mac, allowing us to demonstrate the new technique on different notions. On the other hand, the three constructions are the permutation-

based variants of important block cipher-based schemes LRW2 [27], EDM [16], nEHtM [21], which have already received much attention in the field.

Firstly, we consider the 2-round TEM construction that was proposed by Cogliati et al. [15]. They showed that 2-round TEM achieves 2n/3-bit security in the single-user setting. In Section 5, we apply the public permutation mirror theory suitable for tsprp (Theorem 1 (a)) to the TEM construction, and show that TEM also achieves 2n/3-bit security in the multi-user setting.

Secondly, we consider the pEDM construction that was proposed by Dutta et al. [22]. Again, pEDM was showed to achieve 2n/3-bit security in the single-user setting. In Section 6, we apply the public permutation mirror theory suitable for prf (Theorem 1 (b)) to pEDM construction, and show that pEDM also achieves 2n/3-bit security in the multi-user setting. In this example we can clearly see that the analysis in the multi-user setting is more complex than the one in the single-user setting.

Thirdly, we consider the nEHtM_p MAC algorithm, proposed by Dutta and Nandi [20]. They showed that nEHtM_p based on a single permutation (using domain separation) achieves 2n/3-bit security. We first note that the proof of [20] is incomplete since, according to our framework for MAC designs, the authors missed some bad events in their analysis. This observation was also verified by the authors [10]. Some of these additional bad events, however, require involved arguments to bound. In order to solve this problem, we modify the nEHtM_p construction by adding an extra universal hash function call. This modified construction uses more randomness, which in turn enables us to bound the additional bad events easily. We would like to note that our analysis does not imply infeasibility in fixing the proof of nEHtM_p. In Section 7, we will prove the *multiuser* security of this modified variant nEHtM_p using our public permutation extended mirror theory (Theorem 2 (b)), and we show that it achieves 2n/3-bit security in the multi-user setting.

We believe that the techniques have a wide range of applications in the future design of public permutation based schemes. For example, when building nonce-less MAC algorithms and (authenticated) encryption schemes with beyond birthday bound security using public permutations, as done in the case of block cipher-based mirror theory [13, 28].

2 Preliminaries

For $n \in \mathbb{N}$, we denote by [n] the shorthand notation for $\{1, \ldots, n\}$, and by $\{0, 1\}^n$ the set of bit strings of length n. For two bit strings $X, Y \in \{0, 1\}^n$, we denote their bitwise addition as $X \oplus Y$. For a value Z, we denote by $A \leftarrow Z$ the assignment of Z to the variable A. For a finite set S, we denote by $S \xleftarrow{\$} S$ the uniformly random selection of S from S. We denote by Func(m, n) the set of all functions that map $\{0, 1\}^n$ to $\{0, 1\}^n$, and by Func(n) the set of all functions that $\{0, 1\}^n$. We denote by $\operatorname{Perm}(n)$ the set of all permutations on $\{0, 1\}^n$, and by $\operatorname{Perm}(t, n)$ the set of all functions $\tilde{\pi} : \{0, 1\}^t \times \{0, 1\}^n \to \{0, 1\}^n$

such that $\tilde{\pi}(T, \cdot)$ is in Perm(n) for all $T \in \{0, 1\}^t$. For any integers a, b such that $1 \leq b \leq a$, we have $(a)_b = a \cdot (a-1) \dots (a-b+1)$ and $(a)_0 = 1$.

For $q \in \mathbb{N}$, we denote by x^{*q} the q-tuple (x_1, \ldots, x_q) , and by \hat{x}^{*q} the set $\{x_i : i \in [q]\}$. By an abuse of notation we also use x^{*q} to denote the multiset $\{x_i : i \in [q]\}$, and we denote by $\mu(x^{*q}, x')$ the multiplicity of $x' \in x^{*q}$. For two disjoint sets P and Q, we denote their (disjoint) union as $P \sqcup Q$.

2.1 Tweakable Block Ciphers Based on Public Permutations

For $k, n, r, t, u \in \mathbb{N}$, consider a tweakable block cipher $\tilde{E} : \{0, 1\}^k \times \{0, 1\}^t \times \{0, 1\}^n \to \{0, 1\}^n$ that is based on $\pi_1, \ldots, \pi_r \stackrel{*}{\leftarrow} \operatorname{Perm}(n)$, such that for fixed key $K \in \{0, 1\}^k$, the function $\tilde{E}_K(T, \cdot) = \tilde{E}(K, T, \cdot)$ is a permutation on $\{0, 1\}^n$. We denote its inverse (for fixed key and tweak) by $\tilde{E}_K^{-1}(T, \cdot) = \tilde{E}^{-1}(K, T, \cdot)$, and \tilde{E}_K^{-1} should behave independently for different tweaks. We will consider the multi-user tweakable strong pseudorandom permutation (mu-tsprp) security of \tilde{E} , where the distinguisher \mathcal{D} is given two-directional access to either $(\tilde{E}_{K_1}^{\pm}, \ldots, \tilde{E}_{K_u}^{\pm}, \pi_1^{\pm}, \ldots, \pi_r^{\pm})$ for secret keys $K_1, \ldots, K_u \stackrel{*}{\leftarrow} \{0, 1\}^k$, or $(\tilde{\pi}_1^{\pm}, \ldots, \tilde{\pi}_u^{\pm}, \pi_1^{\pm}, \ldots, \tilde{\pi}_u^{\pm}, \tilde{P}\operatorname{erm}(t, n)$. The goal is to determine which world it interacted with:

$$\mathbf{Adv}_{\tilde{E}}^{\mathrm{mu-tsprp}}(\mathcal{D}) = \left| \Pr\left[\mathcal{D}^{\tilde{E}_{K_{1}}^{\pm},\dots,\tilde{E}_{K_{u}}^{\pm},\pi_{1}^{\pm},\dots,\pi_{r}^{\pm}} = 1 \right] - \Pr\left[\mathcal{D}^{\tilde{\pi}_{1}^{\pm},\dots,\tilde{\pi}_{u}^{\pm},\pi_{1}^{\pm},\dots,\pi_{r}^{\pm}} = 1 \right] \right|$$

Here the superscript \pm indicates that the distinguisher has bi-directional access. When u = 1, we consider the single-user security of \tilde{E} , and we simply denote \mathcal{D} 's advantage in distinguishing the real world from random by $\mathbf{Adv}_{\tilde{E}}^{\mathrm{tsprp}}(\mathcal{D})$.

2.2 Pseudorandom Functions Based on Public Permutations

For $k, m, n, r, u \in \mathbb{N}$, consider a pseudorandom function $F: \{0, 1\}^k \times \{0, 1\}^m \rightarrow \{0, 1\}^n$ that is based on $\pi_1, \ldots, \pi_r \stackrel{\text{s}}{\leftarrow} \operatorname{Perm}(n)$, such that for fixed key $K \in \{0, 1\}^k, F_K(\cdot) = F(K, \cdot)$ is a function that maps $\{0, 1\}^m$ to $\{0, 1\}^n$. We will consider the multi-user pseudorandom function (mu-prf) security of F, where the distinguisher \mathcal{D} is given access to either $(F_{K_1}, \ldots, F_{K_u}, \pi_1^{\pm}, \ldots, \pi_r^{\pm})$ for secret keys $K_1, \ldots, K_u \stackrel{\text{s}}{\leftarrow} \{0, 1\}^k$, or $(\varphi_1, \ldots, \varphi_u, \pi_1^{\pm}, \ldots, \pi_r^{\pm})$ for $\varphi_1, \ldots, \varphi_u \stackrel{\text{s}}{\leftarrow}$ Func(n). The goal is to determine which world it interacted with:

$$\mathbf{Adv}_{F}^{\mathrm{mu-prf}}(\mathcal{D}) = \left| \Pr\left[\mathcal{D}^{F_{K_{1}},\dots,F_{K_{u}},\pi_{1}^{\pm},\dots,\pi_{r}^{\pm}} = 1 \right] - \Pr\left[\mathcal{D}^{\varphi_{1},\dots,\varphi_{u},\pi_{1}^{\pm},\dots,\pi_{r}^{\pm}} = 1 \right] \right| \,.$$

Here the superscript \pm for π 's indicates that the distinguisher has bi-directional access. When u = 1, we consider the single-user security of F, and we simply denote \mathcal{D} 's advantage in distinguishing the real world from random by $\mathbf{Adv}_{F}^{\mathrm{prf}}(\mathcal{D})$.

2.3 Nonce-Based MAC Algorithms Based on Public Permutations

For $k, n, r, t \in \mathbb{N}$, consider a nonce-based message authentication code (MAC) algorithm $F: \{0,1\}^k \times \{0,1\}^n \times \{0,1\}^* \to \{0,1\}^t$ that is based on $\pi_1, \ldots, \pi_r \stackrel{\$}{\leftarrow}$

Perm(n). For any fixed key $K \in \{0,1\}^k$, we write $F_K(\cdot, \cdot) = F(K, \cdot, \cdot)$. We denote by Ver: $\{0,1\}^k \times \{0,1\}^n \times \{0,1\}^* \times \{0,1\}^t \to 1/0$ the verification oracle that is based on π_1, \ldots, π_r , which takes as input a key $K \in \{0,1\}^k$, a nonce $N \in \{0,1\}^n$, a message $M \in \{0,1\}^*$, and tag $T \in \{0,1\}^t$, and outputs 1 if the tag T is correct and 0 otherwise.

For $u \in \mathbb{N}$, the multi-user message authentication code (mu-mac) security of F is measured by considering a distinguisher \mathcal{D} that is given access to $(F_{K_1}, \operatorname{Ver}_{K_1}), \ldots, (F_{K_u}, \operatorname{Ver}_{K_u})$ for secret keys $K_1, \ldots, K_u \stackrel{\$}{\leftarrow} \{0, 1\}^k$, and the primitive oracles π_1, \ldots, π_r . For any $j = 1, \ldots, u$, the goal of \mathcal{D} is to fool the verification oracle with a valid but new (j, N, M, T), and its advantage with respect to this task is defined as

$$\mathbf{Adv}_{F}^{\mathrm{mu-mac}}(\mathcal{D}) = \Pr\left[K_{1}, \dots, K_{u} \xleftarrow{\hspace{0.1cm}} \{0, 1\}^{k}:\right.$$
$$\mathcal{D}^{(F_{K_{1}}, \mathsf{Ver}_{K_{1}}), \dots, (F_{K_{u}}, \mathsf{Ver}_{K_{u}}), \pi_{1}^{\pm}, \dots, \pi_{r}^{\pm} \text{ forges}\right],$$

where "forges" means that the distinguisher enters a tuple (j, N, M, T) such that $\operatorname{Ver}_{K_i}(N, M, T)$ returns 1 and $F_{K_i}(N, M)$ has never been queried.

We call a MAC query to the *j*-th user (j, N, M) a faulty query if the distinguisher \mathcal{D} has already queried F_{K_j} with the same nonce N and a different message M. The distinguisher \mathcal{D} is allowed to make at most μ faulty MAC queries over u users. We call \mathcal{D} a nonce-respecting adversary if $\mu = 0$, and nonce-misusing if $\mu \geq 1$. We stress that \mathcal{D} may always repeat nonces in its verification queries.

It will be more convenient to express $\operatorname{Adv}_{F}^{\operatorname{mu-mac}}(\mathcal{D})$ as a distinguisher's advantage. For $j = 1, \ldots, u$, we define perfectly random oracles $\operatorname{Rand}_{j}: \{0, 1\}^{n} \times \{0, 1\}^{t} \to \{0, 1\}^{t}$, and rejection oracles $\operatorname{Rej}_{j}: \{0, 1\}^{n} \times \{0, 1\}^{t} \times \{0, 1\}^{t} \to 0$.

To obtain an upper bound for the forging advantage of a message authentication code F with respect to the distinguisher \mathcal{D} , we consider another distinguisher \mathcal{D}' , that is given access to either the real world oracles $\mathcal{O}, \pi_1^{\pm}, \ldots, \pi_r^{\pm}$, or the ideal world oracles $\mathcal{P}, \pi_1^{\pm}, \ldots, \pi_r^{\pm}$. Then, \mathcal{D}' 's advantage is upper bounded by:

$$\mathbf{Adv}_{F}^{\mathrm{mu-mac}}(\mathcal{D}') \leq \left| \Pr\left[\mathcal{D}'^{\mathcal{O}, \pi_{1}^{\pm}, \dots, \pi_{r}^{\pm}} = 1 \right] - \Pr\left[\mathcal{D}'^{\mathcal{P}, \pi_{1}^{\pm}, \dots, \pi_{r}^{\pm}} = 1 \right] \right|,$$

with $\mathcal{O} = ((F_{K_1}, \mathsf{Ver}_{K_1}), \dots, (F_{K_u}, \mathsf{Ver}_{K_u}))$ for secret keys $K_1, \dots, K_u \stackrel{\hspace{0.1em}\mathsf{\scriptscriptstyle\$}}{\leftarrow} \{0, 1\}^k$, and $\mathcal{P} = ((\mathsf{Rand}_1, \mathsf{Rej}_1), \dots, (\mathsf{Rand}_u, \mathsf{Rej}_u)).$

Here the superscript \pm for the π_i 's indicates that the distinguisher has bidirectional access. We call a distinguisher \mathcal{D}' non-trivial if it never makes a query (j, N, M, T) to its *j*-th verification oracle when a previous query (j, N, M) to its *j*-th MAC oracle returned *T*. When u = 1, we consider the single-user security of *F*, and we simply denote \mathcal{D}' 's advantage in distinguishing the real world from random by $\mathbf{Adv}_F^{\mathrm{mac}}(\mathcal{D}')$.

2.4 Universal Hash Functions

For $n \in \mathbb{N}$, let $H: \mathcal{K}_h \times \mathcal{M} \to \{0, 1\}^n$ such that for $K_h \in \mathcal{K}_h, H_{K_h}(\cdot) = H(K_h, \cdot)$ is called an ϵ -almost XOR universal (ϵ -AXU) hash function [29] if for all distinct $M, M' \in \mathcal{M}$ and all $C \in \{0, 1\}^n$, we have

$$\Pr\left[K_h \stackrel{*}{\leftarrow} \mathcal{K}_h \colon H_{K_h}(M) \oplus H_{K_h}(M') = C\right] \leq \epsilon \,.$$

For $q \in \mathbb{N}$, fix $M_1, \ldots, M_q \in \mathcal{M}$. For $K_h \in \mathcal{K}_h$, let $X_i = H_{K_h}(M_i)$ for $i = 1, \ldots, q$. We define an equivalence relation \sim on [q] as: $\alpha \sim \beta$ if and only if $X_\alpha = X_\beta$, for $\alpha, \beta \in [q]$. For $r \in \mathbb{N}$, we denote by $\mathcal{P}_1, \ldots, \mathcal{P}_r$ the non-trivial equivalence classes of [q] with respect to \sim , and we define $\nu_i = |\mathcal{P}_i| \geq 2$ for $i = 1, \ldots, r$. Jha and Nandi [27] proved the following lemmas, that will be useful in our security proof.

Lemma 1. Let ν_i , i = 1, ..., r, be the random variables as defined above. Then, we have

$$oldsymbol{E}\Big[\sum_{i=1}^r
u_i\Big] \leq q^2 \epsilon/2\,, \qquad oldsymbol{E}\Big[\sum_{i=1}^r
u_i^2\Big] \leq 2q^2 \epsilon\,.$$

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

$$\Pr[\nu_{\max} \ge a] \le \frac{2q^2\epsilon}{a^2}.$$

2.5 Expectation Method

In this work, we use the expectation method by Hoang and Tessaro [26], a generalization of Patarin's H-coefficient technique [9,33].

Consider two oracles \mathcal{O} and \mathcal{P} , and a deterministic distinguisher \mathcal{D} that has query access to either of these oracles. The distinguisher's goal is to distinguish both worlds, and we denote by

$$\mathbf{Adv}(\mathcal{D}) = \left| \Pr \left[\mathcal{D}^{\mathcal{O}} = 1 \right] - \Pr \left[\mathcal{D}^{\mathcal{P}} = 1 \right] \right|$$

its advantage. We define a transcript τ which summarizes all query-response tuples learned by \mathcal{D} during its interaction with its oracle \mathcal{O} or \mathcal{P} . We denote by $X_{\mathcal{O}}$ and $X_{\mathcal{P}}$ the random variables equal to transcript produced when interacting with \mathcal{O} and \mathcal{P} , respectively. We call a transcript $\tau \in \mathcal{T}$ attainable if $\Pr[X_{\mathcal{P}} = \tau] > 0$, or in other words if the transcript τ can be obtained from an interaction with \mathcal{P} .

Lemma 3 (expectation method [26]). Consider a deterministic distinguisher \mathcal{D} . Define a partition $\mathcal{T} = \mathcal{T}_{good} \sqcup \mathcal{T}_{bad}$, where \mathcal{T}_{good} is the subset of \mathcal{T} which contains all the "good" transcripts and \mathcal{T}_{bad} is the subset with all the "bad" transcripts. Let $\phi: \mathcal{T} \to [0, \infty)$ be a non-negative function mapping any attainable transcript to a non-negative real value, such that for all $\tau \in \mathcal{T}_{good}$:

$$\frac{\Pr[X_{\mathcal{O}} = \tau]}{\Pr[X_{\mathcal{P}} = \tau]} \ge 1 - \phi(\tau) \,. \tag{1}$$

Then, we have $\operatorname{Adv}(\mathcal{D}) \leq E[\phi(X_{\mathcal{P}})] + \Pr[X_{\mathcal{P}} \in \mathcal{T}_{\operatorname{bad}}].$

The H-coefficients technique can be seen as a simple corollary of the expectation method when ϕ is equal to a constant function.

Preliminary Observations. For $\pi \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$ and a permutation queries transcript τ_{π} , we say that π extends τ_{π} , denoted $\pi \vdash \tau_{\pi}$, if $\pi(u) = v$ for all $(u, v) \in \tau_{\pi}$. By extension, for $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_r) \stackrel{\$}{\leftarrow} \left(\operatorname{Perm}(n)\right)^r$ and a tuple of permutation queries transcript $\tau_{\pi} = (\tau_{\pi_1}, \ldots, \tau_{\pi_r})$, we say that $\boldsymbol{\pi}$ extends τ_{π} , denoted $\boldsymbol{\pi} \vdash \tau_{\pi}$, if $\pi_i \vdash \tau_{\pi_i}$ for $i = 1, \ldots, r$.

Consider an attainable transcript $\tau \in \mathcal{T}_{good}$, and let \mathcal{P} be an uniformly chosen random oracle. For permutation based constructions, let $\tau = (\tau_0, \tau_{\pi})$, where τ_0 contains queries to the construction oracle \mathcal{O} or \mathcal{P} , and τ_{π} contains the queries to the primitive oracles $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_r)$. To compute $\Pr[X_{\mathcal{O}} = \tau]$ and $\Pr[X_{\mathcal{P}} = \tau]$, it suffices to compute the probability of oracles that could result in view τ . We first consider the ideal world oracle \mathcal{P} , and obtain

$$\Pr[X_{\mathcal{P}} = \tau] = \frac{1}{|\mathcal{K}|^r} \cdot \left(\frac{1}{(2^n)_p}\right)^r \cdot \Pr[\mathcal{P} \colon \mathcal{P} \vdash \tau_0]$$

The first term corresponds to the number of dummy keys that are drawn uniformly at random; the second term is the probability that π extends τ_{π} ; and the last term is the probability that \mathcal{P} extends τ_0 .

Similarly we say that a real world oracle \mathcal{O} extends τ if it extends τ_0 and τ_{π} . For $K \stackrel{s}{\leftarrow} \mathcal{K}^r$, we have

$$\Pr[X_{\mathcal{O}} = \tau] = \frac{1}{|\mathcal{K}|^r} \cdot \left(\frac{1}{(2^n)_p}\right)^r \cdot \Pr\left[\boldsymbol{\pi} \xleftarrow{} \left(\operatorname{Perm}(n)\right)^r : \mathcal{O}_K^{\boldsymbol{\pi}} \vdash \tau_0 \mid \boldsymbol{\pi} \vdash \tau_{\boldsymbol{\pi}}\right].$$

The first term corresponds to the number of randomly drawn keys that are used in the construction; the second term is the probability that π extends τ_{π} ; and the last term is the probability that \mathcal{O}_{K}^{π} extends τ_{0} , given that π extends τ_{π} .

the last term is the probability that \mathcal{O}_{K}^{π} extends τ_{0} , given that π extends τ_{π} . Let $\rho(\tau) = \Pr\left[\pi \stackrel{\$}{\leftarrow} \left(\operatorname{Perm}(n)\right)^{r} : \mathcal{O}_{K}^{\pi} \vdash \tau_{0} \mid \pi \vdash \tau_{\pi}\right]$. Take for instance r = 2, and assume that each primitive query transcript contains p queries to the given permutation. Suppose we sample distinct outputs of π_{1} (resp., π_{2}) over for example q_{V} (resp., q_{Y}) distinct inputs. Then, it is easy to see that $\rho(\tau) = h_{q}/(2^{n} - p)_{q_{V}-p}(2^{n} - p)_{q_{Y}-p}$, where h_{q} is the number of solutions of distinct outputs of π_{1} and π_{2} . Then we have

$$\frac{\Pr[X_{\mathcal{O}} = \tau]}{\Pr[X_{\mathcal{P}} = \tau]} = \rho(\tau) / \Pr[\mathcal{P} \colon \mathcal{P} \vdash \tau_0] \ge 1 - \varepsilon_1 \,. \tag{2}$$

3 (Extended) Mirror Theory in Ideal Permutation Model

We explain the general settings behind the traditional (extended) mirror theory in Section 3.1, which involves only bi-variate affine equations and possible non-equations. In Section 3.2, we introduce the new public permutation mirror theory, that takes primitive queries into account by including uni-variate affine equations.

3.1 System of Bi-Variate Affine Equations and Non-Equations

Let $q_m, q_a, q_V, q_Y \ge 1$. Let $\mathcal{V} = \{v_1, \ldots, v_{q_V}\}$ be a set of q_V unknowns and $\mathcal{Y} = \{y_1, \ldots, y_{q_Y}\}$ be a set of q_Y unknowns. We consider a system \mathcal{E}_m of q_m bi-variate affine equations

$$\mathcal{E}_m = \{ v_{I_1} \oplus y_{I_1} = \lambda_1, \dots, v_{I_{q_m}} \oplus y_{I_{q_m}} = \lambda_{q_m} \}.$$

In some cases (for example mac security), we also need to consider a system \mathcal{E}_a of q_a bi-variate affine non-equations

$$\mathcal{E}_a = \{ v'_{J_1} \oplus y'_{J_1} \neq \lambda'_1, \dots, v'_{J_{q_a}} \oplus y'_{J_{q_a}} \neq \lambda'_{q_a} \},\$$

where v_{I_i} 's, y_{I_i} 's, v'_{J_j} 's, and y'_{J_j} 's are unknowns, and λ_i 's and λ'_j 's are knowns, for $i = 1, \ldots, q_m$ and $j = 1, \ldots, q_a$. We want to state that the sets \mathcal{V} and \mathcal{Y} are disjoint.

We define two surjective index mappings:

$$\varphi_{V}: \{I_{1}, \dots, I_{q_{m}}, J_{1}, \dots, J_{q_{a}}\} \to \{1, \dots, q_{V}\}, \varphi_{Y}: \{I_{1}, \dots, I_{q_{m}}, J_{1}, \dots, J_{q_{a}}\} \to \{1, \dots, q_{Y}\}.$$

such that $q_V, q_Y \leq q_m + q_a$. Note that I_i and J_j are respectively the indices of the unknowns in \mathcal{E}_m and \mathcal{E}_a . However, multiple unknowns with different indices can be the same. In that case, these unknown are all mapped to the same value using φ_V or φ_Y . The system $\mathcal{E} = \mathcal{E}_m \sqcup \mathcal{E}_a$ is uniquely determined by $(\varphi'_V, \varphi'_Y, \lambda^{*q_m}, \lambda'^{*q_a}).$

Consider a graph $\mathcal{G}(\mathcal{E}) = (\mathcal{V}, \mathcal{Y}, \mathcal{S} \sqcup \mathcal{S}')$, where the edge set is partitioned into two disjoint sets \mathcal{S} and \mathcal{S}' . Here \mathcal{S} and \mathcal{S}' denote the set of λ -labeled edges and the set of λ' -labeled edges, respectively. The graph $\mathcal{G}(\mathcal{E})$ can be seen as a superposition of two subgraphs $\mathcal{G}(\mathcal{E}_m) = (\mathcal{V}, \mathcal{Y}, \mathcal{S})$ and $\mathcal{G}(\mathcal{E}_a) = (\mathcal{V}, \mathcal{Y}, \mathcal{S}')$. Let $\overline{v_s y_t} \in \mathcal{S}$ be an edge for $v_s \in \mathcal{V}$ and $y_t \in \mathcal{Y}$, then $\overline{v_s y_t}$ is labeled with an element in λ^{*q_m} . If the given edge is labeled with λ_i (for $i = 1, \ldots, q_m$), then this edge and the connected vertices v_s and y_t represent the equation $v_s \oplus y_t = \lambda_i$, where $s = \varphi_V(I_i)$ and $t = \varphi_Y(I_i)$. Similarly, let $\overline{v_s y_t} \in \mathcal{S}'$ be an edge for $v_s \in \mathcal{V}$ and $y_t \in \mathcal{Y}$, then $\overline{v_s y_t}$ is labeled with an element in λ'^{*q_a} . If the given edge is labeled with λ'_i (for $j = 1, \ldots, q_a$), then this edge and the connected vertices v_s and y_t represent the non-equation $v_s \oplus y_t \neq \lambda'_j$, where $s = \varphi_V(J_j)$ and $t = \varphi_Y(J_j)$. Here, each equation in \mathcal{E}_m corresponds to a unique λ -labeled edge in $\mathcal{G}(\mathcal{E}_m)$, and each non-equation in \mathcal{E}_a corresponds to a unique λ' -labeled edge in $\mathcal{G}(\mathcal{E}_a)$. Note that when the system of non-equations \mathcal{E}_a is empty, then the graph $\mathcal{G}(\mathcal{E})$ does not contain isolated vertices, every vertex is incident with at least one λ -labeled edge. Otherwise, the subgraph $\mathcal{G}(\mathcal{E}_m)$ may contain isolated vertices, and these vertices are connected with a λ' -labeled edge in $\mathcal{G}(\mathcal{E}_a)$.

We say two distinct equations in \mathcal{E}_m are in the same component if and only if the corresponding edges (or vertices) in $\mathcal{G}(\mathcal{E}_m)$ are in the same component. Let $\ell > 0$ and a path

$$\mathcal{L}: a_0 \stackrel{\lambda_1}{-} a_1 \stackrel{\lambda_2}{-} \dots \stackrel{\lambda_\ell}{-} a_\ell$$

in $\mathcal{G}(\mathcal{E}_m)$, for some vertices $a_0, a_1, \ldots, a_\ell \in \mathcal{V} \sqcup \mathcal{Y}$ that are in the same component. The label of \mathcal{L} is defined as

$$\lambda(\mathcal{L}) = \lambda_1 \oplus \lambda_2 \oplus \ldots \oplus \lambda_\ell$$

In case \mathcal{E}_a is empty, a graph $\mathcal{G}(\mathcal{E})$ is called a *good* graph if its subgraph $\mathcal{G}(\mathcal{E}_m)$ satisfies the following properties.

Definition 1 (acyclic). There is an unique path \mathcal{L} in the subgraph $\mathcal{G}(\mathcal{E}_m)$ between any two vertices a and b in the same connected component, for $a, b \in \mathcal{V} \sqcup \mathcal{Y}$.

Definition 2 (non-degeneracy). For all paths \mathcal{L} of even length at least 2 in the subgraph $\mathcal{G}(\mathcal{E}_m)$, we have $\lambda(\mathcal{L}) \neq 0$.

Definition 3 (ξ **-block-maximality).** For a component \mathcal{I} , we denote its size by $\xi(\mathcal{I})$, which is the number of vertices in \mathcal{I} . We denote the maximum component size by ξ_{\max} such that $\xi(\mathcal{I}) \leq \xi_{\max}$ for all \mathcal{I} in $\mathcal{G}(\mathcal{E}_m)$.

In case the system \mathcal{E}_a contains at least one non-equation, a graph $\mathcal{G}(\mathcal{E})$ is called a good graph if its subgraph $\mathcal{G}(\mathcal{E}_m)$ satisfies the above three properties, and if $\mathcal{G}(\mathcal{E})$ also satisfies the following property.

Definition 4 (non-zero cycle label (NCL)). If vertices v and y are connected with a λ' -labeled edge, then they are not connected by a $\lambda(\mathcal{L})$ -labeled path in $\mathcal{G}(\mathcal{E}_m)$ such that $\lambda(\mathcal{L}) = \lambda'$, for $v \in \mathcal{V}'$ and $y \in \mathcal{Y}'$.

In an edge-labeled bipartite graph \mathcal{G} , we call a component \mathcal{I} of \mathcal{G} an isolated component if \mathcal{I} only contains a path of length one. So \mathcal{I} only contains an edge (v, y, λ) where both v and y have degree 1. We call a component \mathcal{I} of \mathcal{G} a star component if $\xi(\mathcal{I}) \geq 3$, and if there is an unique vertex in \mathcal{I} with degree $\xi(\mathcal{I}) - 1$. We call this vertex the center of \mathcal{I} . Further, we call \mathcal{I} a v- \star (resp., y- \star) component if its center lies in v (resp., y).

3.2 System of Bi-Variate and Uni-Variate Affine Equations and Bi-Variate Affine Non-Equations

In order to handle the primitive queries, we extend the system of bi-variate affine equations and non-equations with 2p uni-variate affine equations. Each uni-variate affine equation defines one primitive queries. Let $q_m, q_a, q_V^p, q_Y^p, p \ge 1$. Let $\mathcal{V}^p = \{v_1, \ldots, v_{q_V^p}\}$ be a set of q_V^p unknowns and $\mathcal{Y}^p = \{y_1, \ldots, y_{q_Y^p}\}$ be a set of q_Y^p unknowns and $\mathcal{Y}^p = \{y_1, \ldots, y_{q_Y^p}\}$ be a set of q_Y^p unknowns and $\mathcal{Y}^p = \{y_1, \ldots, y_{q_Y^p}\}$ be a set of q_Y^p unknowns. The new systems are \mathcal{E}_m^p , that contains q_m bi-variate affine equations and 2p uni-variate affine equations

$$\mathcal{E}_{m}^{p} = \{ v_{I_{1}} \oplus y_{I_{1}} = \lambda_{1}, \dots, v_{I_{q_{m}}} \oplus y_{I_{q_{m}}} = \lambda_{q_{m}}, \\ v_{I_{q_{m}+1}} = \lambda_{q_{m}+1}, \dots, v_{I_{q_{m}+p}} = \lambda_{q_{m}+p}, \\ y_{I_{q_{m}+1}} = \lambda_{q_{m}+p+1}, \dots, y_{I_{q_{m}+p}} = \lambda_{q_{m}+2p} \}.$$

In some cases (for example mac security), we also need to consider a system \mathcal{E}_a of q_a bi-variate affine non-equations

$$\mathcal{E}_a = \{ v'_{J_1} \oplus y'_{J_1} \neq \lambda'_1, \dots, v'_{J_{q_a}} \oplus y'_{J_{q_a}} \neq \lambda'_{q_a} \},\$$

where v_{I_i} 's, y_{I_i} 's, v'_{J_j} 's, and y'_{J_j} 's are unknowns, and λ_k 's and λ'_j 's are knowns, for $i = 1, \ldots, q_m + p$, $j = 1, \ldots, q_a$, and $k = 1, \ldots, q_m + 2p$, such that $\lambda_{q_m+1} \neq \ldots \neq \lambda_{q_m+p}$ and $\lambda_{q_m+p+1} \neq \ldots \neq \lambda_{q_m+2p}$. We want to state that the sets \mathcal{V}^p and \mathcal{Y}^p are disjoint.

We define two surjective index mappings:

$$\varphi_V^p \colon \{I_1, \dots, I_{q_m+p}, J_1, \dots, J_{q_a}\} \to \{1, \dots, q_V^p\}, \varphi_V^p \colon \{I_1, \dots, I_{q_m+p}, J_1, \dots, J_{q_a}\} \to \{1, \dots, q_V^p\},$$

such that $q_V, q_Y \leq q_m + q_a + p$. The system $\mathcal{E}^p = \mathcal{E}^p_m \sqcup \mathcal{E}_a$ is uniquely determined by $(\varphi^p_V, \varphi^p_Y, \lambda^{*q_m+2p}, \lambda'^{*q_a})$.

Since the last 2p uni-variate affine equations define the values of the 2p unknowns v_{I_i} 's and y_{I_i} 's exactly, for $i = q_m + 1, \ldots, q_m + p$. Hence, we know that exactly p unknowns in \mathcal{V}^p and p unknowns in \mathcal{Y}^p are already well defined by the system \mathcal{E}_m^p . We define

$$V_0 = \{ v_{\varphi_V^p(I_{q_m+1})}, \dots, v_{\varphi_V^p(I_{q_m+p})} \}, \qquad Y_0 = \{ y_{\varphi_V^p(I_{q_m+1})}, \dots, y_{\varphi_V^p(I_{q_m+p})} \},$$

as the sets that contain these 2p unknowns such that $|V_0| = p$ and $|Y_0| = p$. We are particularly interested in the unknowns from the sets $\mathcal{V}^p \setminus V_0$ and $\mathcal{Y}^p \setminus Y_0$, since these are the unknowns that appear in the q_m bi-variate affine equations and q_a bi-variate affine non-equations.

Consider a bipartite edge-labeled graph $\mathcal{G}(\mathcal{E}^p) = (\mathcal{V}^p, \mathcal{Y}^p, \mathcal{S} \sqcup \mathcal{S}')$, the edge set is partitioned into two disjoint sets \mathcal{S} and \mathcal{S}' as before. The graph $\mathcal{G}(\mathcal{E}^p)$ can be seen as a superposition of two subgraphs $\mathcal{G}(\mathcal{E}^p_m) = (\mathcal{V}^p, \mathcal{Y}^p, \mathcal{S})$ and $\mathcal{G}(\mathcal{E}_a) = (\mathcal{V}^p, \mathcal{Y}^p, \mathcal{S}')$. Here, each of the q_m bi-variate affine equations in \mathcal{E}^p_m corresponds to a unique λ -labeled edge in $\mathcal{G}(\mathcal{E}^p_m)$, each non-equation in \mathcal{E}_a corresponds to a unique λ' -labeled edge in $\mathcal{G}(\mathcal{E}^p_m)$, each non-equation in \mathcal{E}_a corresponds to a unique λ' -labeled edge in $\mathcal{G}(\mathcal{E}_a)$, and each of the 2p uni-variate affine equations in \mathcal{E}^p_m corresponds to a vertex with well defined value in $\mathcal{G}(\mathcal{E}^p)$. Note that the subgraph $\mathcal{G}(\mathcal{E}^p_m)$ may contain isolated vertices, and these vertices are either connected with a λ' -labeled edge in $\mathcal{G}(\mathcal{E}_a)$ or they are isolated colliding vertices in $\mathcal{G}(\mathcal{E}^p)$ with a well-defined value. The subgraph $\mathcal{G}(\mathcal{E}^p_m)$ may also contain components that contain vertices with well defined value. We call these components the "colliding components", and the vertices with well defined values the "colliding vertices".

We distinguish two different cases. In the first case, assume that \mathcal{E}_a is empty, hence we will focus on a graph $\mathcal{G}(\mathcal{E}^p)$ where its subgraph $\mathcal{G}(\mathcal{E}^p_m)$ satisfies the (i) acyclic, (ii) non-degeneracy, and (iii) ξ -block-maximality properties (Definition 1-3). In addition, $\mathcal{G}(\mathcal{E}^p_m)$ also needs to satisfy the following property.

Definition 5 (single colliding vertex (SCV)). Each component in the graph $\mathcal{G}(\mathcal{E}_m^p)$ contains at most one colliding vertex.

Note that Definition 5 is necessary in order to give a unique assignment to every vertex in the graph, since if a vertex is assigned with any value, then the labeled edges determine the values of all other vertices in the component containing this vertex. We call any graph that satisfies these four properties a good graph. Given a good graph $\mathcal{G}(\mathcal{E}_m^p) = (\mathcal{V}^p, \mathcal{Y}^p, \mathcal{S})$, a solution to $\mathcal{G}(\mathcal{E}_m^p)$ is an assignment of distinct values to the v vertices in \mathcal{V}^p and distinct values to the yvertices in \mathcal{Y}^p satisfying all λ -labeled equations.

We consider a system of bi-variate and uni-variate affine equations \mathcal{E}_m^p , such that each component in $\mathcal{G}(\mathcal{E}_m^p)$ is either an isolated edge, a star, or isolated colliding vertex. In order to find the number of solutions to $\mathcal{G}(\mathcal{E}_m^p)$, we first decompose the graph $\mathcal{G}(\mathcal{E}_m^p)$ into its connected components such that $\mathcal{G}(\mathcal{E}_m^p) = \mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B} \sqcup \mathcal{C}$, where

$$\mathcal{A} = \mathcal{A}_1 \sqcup \cdots \sqcup \mathcal{A}_{c_1} \sqcup \mathcal{A}_{c_1+1} \sqcup \cdots \sqcup \mathcal{A}_{c_1+c_2} ,$$

$$\mathcal{B} = \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_{c_3} \sqcup \mathcal{B}_{c_3+1} \sqcup \cdots \sqcup \mathcal{B}_{c_3+c_4} ,$$

$$\mathcal{C} = \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_{c_5} ,$$

for some $c_1, c_2, c_3, c_4, c_5 \geq 0$. Here \mathcal{I} is the union of isolated colliding vertices. \mathcal{A} is the union of colliding components, where $\mathcal{A}_1 \sqcup \cdots \sqcup \mathcal{A}_{c_1}$ is the union of colliding components with a colliding v vertex; and $\mathcal{A}_{c_1+1} \sqcup \cdots \sqcup \mathcal{A}_{c_1+c_2}$ is the union of colliding components with a colliding y vertex. \mathcal{B} is the union of the remaining star components (that are not colliding components), where $\mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_{c_3}$ is the union of v- \star components, and $\mathcal{B}_{c_3+1} \sqcup \cdots \sqcup \mathcal{B}_{c_3+c_4}$ is the union of y- \star components. \mathcal{C} is the union of the remaining isolated components (that are not colliding components).

Let q_1, q_2, q_3, q_4 , and q_5 denote the number of equations (edges) in $\mathcal{A}_1 \sqcup \cdots \sqcup \mathcal{A}_{c_1}, \mathcal{A}_{c_1+1} \sqcup \cdots \sqcup \mathcal{A}_{c_1+c_2}, \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_{c_3}, \mathcal{B}_{c_3+1} \sqcup \cdots \sqcup \mathcal{B}_{c_3+c_4}$, and \mathcal{C} , respectively. Therefore, we have $c_5 = q_5$. Note that the equations in \mathcal{E}_m^p can be arranged in any arbitrary order without affecting the number of solutions. For the sake of simplicity, we fix the ordering in such a way that the union \mathcal{I} comes first, followed by \mathcal{A}, \mathcal{B} , and \mathcal{C} . Now, our goal is to give a lower bound on the number of solutions of \mathcal{E}_m^p .

Theorem 1. For positive integers q_m and p, let $\mathcal{G}(\mathcal{E}_m^p) = (\mathcal{V}^p, \mathcal{Y}^p, \mathcal{S})$ be a good graph as described above such that $|\mathcal{S}| = q_m$. Assume that $p + q_m \leq 2^{n-2}$ and $\xi_{\max} \cdot (p+q_m) \leq 2^{n-1}$, let $h(\mathcal{G}(\mathcal{E}_m^p))$ denote the number of solutions to $\mathcal{G}(\mathcal{E}_m^p)$.

(a) For settings such as (t)sprp, we have

$$\frac{h(\mathcal{G}(\mathcal{E}_m^p))\prod_{\lambda'\in\hat{\lambda}^{q_m}}(2^n)_{\mu(\lambda^{*q_m},\lambda')}}{(2^n-p)_{q_2+c_3+q_4+q_5}(2^n-p)_{q_1+q_3+c_4+q_5}} \ge 1 - \frac{\sum_{i=1}^{c_1+c_2}q_m}{2^n} - \frac{3q_m^3}{2^{2n}} - \frac{2(p+q_m)^2}{2^{2n}}\left(\eta+q_m\right).$$

(b) For settings such as prf, weak prf, (t)ccr, we have

$$\frac{h(\mathcal{G}(\mathcal{E}_m^p))2^{nq_m}}{(2^n-p)_{q_2+c_3+q_4+q_5}(2^n-p)_{q_1+q_3+c_4+q_5}} \ge 1 - \frac{2(p+q_m)^2}{2^{2n}} \left(\eta + q_m\right).$$

where $\eta = \sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} (\eta_{i+1}^2 + \eta_{i+1}), \ \eta_j = \xi_j - 1 \ and \ \xi_j \ denotes \ the \ size \ (number \ of \ vertices) \ of \ the \ j-th \ component, \ for \ all \ j \in [c_1 + c_2 + c_3 + c_4 + c_5].$

Proof. The proof is given in Supplementary Material A.

We will illustrate Theorem 1 (a) to tweakable block ciphers in Section 5, and Theorem 1 (b) to PRFs in Section 6. Looking back at the discussion given towards the end of Section 2.5, one can see the motivation behind the difference between the expressions given in Theorem 1 (a) and Theorem 1 (b).

For the second case, the system \mathcal{E}_a contains at least one non-equation. Here we will focus on a graph $\mathcal{G}(\mathcal{E}^p)$ such that its subgraph $\mathcal{G}(\mathcal{E}^p_m)$ satisfies the (i) acyclic, (ii) non-degeneracy, (iii) ξ -block-maximality, (v) NCL, and (iv) SCV properties (Definition 1-5). In addition, $\mathcal{G}(\mathcal{E})$ also need to satisfy the following property.

Definition 6 (non-zero distance label (NDL)). There are no λ' -labeled edges that connect two vertices v and y from two colliding components \mathcal{I}_1 and \mathcal{I}_2 , where the distance between v and y (defined as $v \oplus y$) is λ' , for $v \in \mathcal{V}'^p$ and $y \in \mathcal{Y}'^p$.

Note that if there is a λ' -labeled non-equations between vertices v and y of two different colliding components, then it means $v \oplus y \neq \lambda'$. However, for any colliding component, the values of all vertices in this component are uniquely defined. If the distance $v \oplus y$ is equal to λ' , then we will have a contradiction with the non-equation. Definition 6 actually excludes this situation. We call any graph that satisfies these six properties a *good* graph. Given a good graph $\mathcal{G}(\mathcal{E}^p) = (\mathcal{V}^p, \mathcal{Y}^p, \mathcal{S} \sqcup \mathcal{S}')$, a solution to $\mathcal{G}(\mathcal{E}^p)$ is an assignment of distinct values to the v vertices in \mathcal{V}^p and distinct values to the y vertices in \mathcal{Y}^p satisfying all λ -labeled equations and λ' -labeled non-equations.

We consider a system \mathcal{E}^p with its corresponding graph $\mathcal{G}(\mathcal{E}^p)$ such that each component in the subgraph $\mathcal{G}(\mathcal{E}^p_m)$ is a star, an isolated edge, or an isolated vertex. In order to find the number of solutions to $\mathcal{G}(\mathcal{E}^p)$, we first decompose the subgraph $\mathcal{G}(\mathcal{E}^p_m)$ into its connected components such that $\mathcal{G}(\mathcal{E}^p_m) = \mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B} \sqcup$ $\mathcal{C} \sqcup \mathcal{D}$, with $\mathcal{I}, \mathcal{A}, \mathcal{B}$, and \mathcal{C} the union of components defined before. Here, \mathcal{D} is the union of isolated vertices in the subgraph $\mathcal{G}(\mathcal{E}^p_m)$ that are not colliding vertices, note that these vertices are connected with λ' -labeled edges in the subgraph $\mathcal{G}(\mathcal{E}_a)$. Let c_6 be the number of such isolated v vertices, and c_7 be the number of such isolated y vertices. Again, we fix the ordering in such a way that the union \mathcal{I} comes first, followed by $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and \mathcal{D} . Now, our goal is to give a lower bound on the number of solutions of \mathcal{E}^p .

Theorem 2. For positive integers q_m , q_a and p, let $\mathcal{G}(\mathcal{E}^p) = (\mathcal{V}'^p, \mathcal{Y}'^p, \mathcal{S} \sqcup \mathcal{S}')$ be a good graph as described above such that $|\mathcal{S}| = q_m$, $|\mathcal{S}'| = q_a$. Assume that $p + q_m \leq 2^{n-2}$ and $\xi_{\max} \cdot (p + q_m) \leq 2^{n-1}$, let $h(\mathcal{G}(\mathcal{E}^p))$ denote the number of solutions to $\mathcal{G}(\mathcal{E}^p)$.

(a) For settings such as (t)sprp with non-equations, we have

$$\frac{h(\mathcal{G}(\mathcal{E}^p))\prod_{\lambda'\in\hat{\lambda}^{q_m}}(2^n)_{\mu(\lambda^{*q_m},\lambda')}}{(2^n-p)_{q_2+c_3+q_4+q_5+c_6}(2^n-p)_{q_1+q_3+c_4+q_5+c_7}} \ge 1 - \frac{\sum_{i=1}^{c_1+c_2}q_m}{2^n} - \frac{3q_m^3}{2^{2n}} - \frac{2(p+q_m)^2}{2^{2n}}\left(\eta+q_m\right) - \frac{2q_a}{2^n}.$$

(b) For settings such as mac (prf with non-equations), we have

$$\frac{h(\mathcal{G}(\mathcal{E}^p))2^{nq_m}}{(2^n-p)_{q_2+c_3+q_4+q_5+c_6}(2^n-p)_{q_1+q_3+c_4+q_5+c_7}} \ge 1 - \frac{2(p+q_m)^2}{2^{2n}} \left(\eta + q_m\right) - \frac{2q_a}{2^n} \cdot \frac{1}{2^{2n}} \left(\eta + q_m\right) - \frac{2q_a}{2^n} \cdot \frac{1}{2^n} \left(\eta + q_m\right) - \frac{2q_a}{2^n} \cdot \frac{1}{2^n} \cdot \frac{1}{2^n} \left(\eta + q_m\right) - \frac{2q_a}{2^n} \cdot \frac{1}{2^n} \cdot \frac{1}$$

where $\eta = \sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} (\eta_{i+1}^2 + \eta_{i+1}), \ \eta_j = \xi_j - 1 \ and \ \xi_j \ denotes \ the \ size \ (number \ of \ vertices) \ of \ the \ j-th \ component, \ for \ all \ j \in [c_1 + c_2 + c_3 + c_4 + c_5].$

Proof. The proof is given in Supplementary Material B.

We will illustrate Theorem 2 (b) to nonce-based MAC algorithms in Section 7.

4 A Framework for Security Proof Using Public Permutation Mirror Theory

The goal of this section is to establish a general framework for (multi-user) security proof using Theorem 1-2. Note that a framework for specific security notions such as sprp, tsprp, prf, mac, ... can be derived directly from this framework. We consider an algorithm F which is built on two independent public permutations with the following special structure.

Let $n, s, t \in \mathbb{N}$, and let $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$. One can consider the generic construction $F^{\pi_1,\pi_2} \colon \mathcal{K} \times \mathcal{I}_1 \times \cdots \times \mathcal{I}_s \to \mathcal{R}_1 \times \cdots \times \mathcal{R}_t$ based on π_1 and π_2 , where \mathcal{K} is the key space, $\mathcal{I}_1 \times \cdots \times \mathcal{I}_s$ are the input spaces, and $\mathcal{R}_1 \times \cdots \times \mathcal{R}_t$ are the output spaces. Note that here F can be a tweakable block cipher, a PRF, a MAC algorithm, etc. In this work, we will focus on algorithms that can be viewed as the xor of the public permutations

$$Z = \pi_1(A) \oplus \pi_2(B) \,,$$

for $\pi_1, \pi_2 \stackrel{*}{\leftarrow} \operatorname{Perm}(n)$. Here A, B, and Z are functions of the secret key K, the inputs I_1, \ldots, I_s , and the outputs R_1, \ldots, R_t . Although there are no strict restrictions for Z, we do require that the equality patterns of A and B satisfy certain conditions. More precisely, equality pattern of B should not depend on the value of $\pi_1(A)$ and vice versa. This is the condition to use the mirror theory based lower bound as formalized in [32].

4.1 General Setting and Transcripts.

Let $u \in \mathbb{N}, K_1, \ldots, K_u \stackrel{\$}{\leftarrow} \mathcal{K}$, and $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$. Consider any distinguisher \mathcal{D} that has access to the oracles: $(\mathcal{O}_m, \mathcal{O}_a, \pi_1^{\pm}, \pi_2^{\pm})$ in the real world or $(\mathcal{P}_m, \mathcal{P}_a, \pi_1^{\pm}, \pi_2^{\pm})$ in the ideal world. Here we have $\mathcal{O}_m = (F_{K_1}^{\pi_1, \pi_2}, \ldots, F_{K_u}^{\pi_1, \pi_2})$, and \mathcal{O}_a is the possible set of verification oracles (for example the case of mac notion in Section 2.3). The oracle \mathcal{P}_m is the idealized version of $(F_{K_1}^{\pi_1, \pi_2}, \ldots, F_{K_u}^{\pi_1, \pi_2})$, which these idealized oracles are depend on the considered security notion (see Section 2.1-2.3 for more details), and \mathcal{P}_a is the possible set of rejection oracles. We require that \mathcal{D} is computationally unbounded and deterministic. For user index $j \in \{1, \ldots, u\}, \mathcal{D}$ makes q_m queries to \mathcal{O}_m or \mathcal{P}_m , and these are summarized in a transcript

$$\tau_m = \{ (j^{(1)}, I_1^{(1)}, \dots, I_s^{(1)}, R_1^{(1)}, \dots, R_t^{(1)}), \dots, \\ (j^{(q_m)}, I_1^{(q_m)}, \dots, I_s^{(q_m)}, R_1^{(q_m)}, \dots, R_t^{(q_m)}) \} ,$$

and q_a queries to \mathcal{O}_a or \mathcal{P}_a , these are summarized in a transcript

$$\tau_a = \{ (j'^{(1)}, I_1'^{(1)}, \dots, I_s'^{(1)}, R_1'^{(1)}, \dots, R_t'^{(1)}, b'^{(1)}), \dots, \\ (j'^{(q_a)}, I_1'^{(q_a)}, \dots, I_s'^{(q_a)}, R_1'^{(q_a)}, \dots, R_t'^{(q_a)}, b'^{(q_a)}) \}.$$

Note that τ_a is empty for notions where no verification oracles are considered (such as sprp, tsprp, prf, tccr, etc). \mathcal{D} also makes p primitive queries to π_1^{\pm} and p primitive queries to π_2^{\pm} , and like before, these are respectively summarized in transcripts τ_1 and τ_2 . We assume that τ_m , τ_a , τ_1 , and τ_2 do not contain duplicate elements. After \mathcal{D} 's interaction with the oracles, but before it outputs its decision, we disclose the keys K_1, \ldots, K_u to the distinguisher. In the real world, these are the keys used in the construction. In the ideal world, K_1, \ldots, K_u are dummy keys that are drawn uniformly at random. The complete view is denoted $\tau = (\tau_m, \tau_a, \tau_1, \tau_2, K_1, \ldots, K_u)$.

4.2 Attainable Index Mappings.

In the real world, each query $(j^{(i)}, I_1^{(i)}, \ldots, I_s^{(i)}, R_1^{(i)}, \ldots, R_t^{(i)}) \in \tau_m$ corresponds to an evaluation of the $j^{(i)}$ -th oracle in \mathcal{O}_m , each query $(j'^{(a)}, I_1'^{(a)}, \ldots, I_s'^{(a)}, R_1'^{(a)}, \ldots, R_t'^{(a)}, b'^{(a)}) \in \tau_a$ corresponds to an evaluation of the $j'^{(a)}$ -th oracle in \mathcal{O}_a , and each primitive query $(u, v) \in \tau_1$ (resp., $(x, y) \in \tau_2$) corresponds to an evaluation of the primitive oracle π_1^{\pm} (resp., π_2^{\pm}). Note that each algorithm F consists of an evaluation of π_1 and an evaluation of π_2 . For the queries in τ_m , these are of the form $A^{(i)} \mapsto \pi_1(A^{(i)})$ and $B^{(i)} \mapsto \pi_2(B^{(i)})$ such that $\pi_1(A^{(i)}) \oplus \pi_2(B^{(i)}) = Z^{(i)}$. Likewise, for the queries in τ_a , there are evaluations $A'^{(a)} \mapsto \pi_1(A'^{(a)})$ and $B'^{(a)} \mapsto \pi_2(B'^{(a)})$, such that $\pi_1(A'^{(a)}) \oplus \pi_2(B'^{(a)}) \neq Z'^{(a)}$. The values of $A^{(i)}, B^{(i)}, Z^{(i)}$ and $A'^{(a)}, B'^{(a)}, Z'^{(a)}$ are specific for the particular construction, and can be deduced from τ . Without loss of generality, we assume that all primitive queries are made in the forward direction, then these are of the form $u \mapsto \pi_1(u)$ or $x \mapsto \pi_2(x)$ such that $\pi_1(u) = v$ and $\pi_2(x) = y$. The transcript τ defines $q_m + 2p$ equations and q_a non-equations on the unknowns, and these (non)-equations are

$$\mathcal{E}_{m}^{p} = \begin{cases} \pi_{1}(A^{(1)}) \oplus \pi_{2}(B^{(1)}) = Z^{(1)}, \\ \vdots \\ \pi_{1}(A^{(q_{m})}) \oplus \pi_{2}(B^{(q_{m})}) = Z^{(q_{m})}, & \mathcal{E}_{a} = \begin{cases} \pi_{1}(A^{\prime(1)}) \oplus \pi_{2}(B^{\prime(1)}) \neq Z^{\prime(1)}, \\ \vdots \\ \pi_{1}(u) = v \quad \text{for } (u, v) \in \tau_{1}, \\ \pi_{2}(x) = y \quad \text{for } (x, y) \in \tau_{2}, \end{cases}$$

$$(3)$$

In the above $q_m + 2p$ equations, some of the unknowns may be equal to each other. We have that $\pi_1(A^{(i)}) \neq \pi_1(A^{(k)})$ if and only if $A^{(i)} \neq A^{(k)}$, and $\pi_2(B^{(i)}) \neq \pi_2(B^{(k)})$ if and only if $B^{(i)} \neq B^{(k)}$. No condition holds for $\pi_1(A^{(i)})$ versus $\pi_2(B^{(i)})$, as these are defined by independent permutations. The same holds for verification queries and primitive queries. However, no a priori condition holds for (non-)equality between values $\pi_1(A^{(i)})$ versus $\pi_1(A'^{(a)})$ versus $\pi_1(u)$, and $\pi_2(B^{(i)})$ versus $\pi_2(B'^{(a)})$ versus $\pi_2(x)$.

Thus,

$$\{\pi_1(A^{(i)})\}_{1 \le i \le q_m} \cup \{\pi_1(A'^{(a)})\}_{1 \le a \le q_a} \cup \{\pi_1(u)\}_{(u,v) \in \tau_1}, \{\pi_2(B^{(i)})\}_{1 \le i \le q_m} \cup \{\pi_2(B'^{(a)})\}_{1 \le a \le q_a} \cup \{\pi_2(x)\}_{(x,y) \in \tau_2},$$

are identified with two sets of unknowns $\mathcal{V}'^p = \{v_1, \ldots, v_{q_{V'}^p}\}$ and $\mathcal{Y}'^p = \{y_1, \ldots, y_{q_{Y'}^p}\}$, with $q_{V'}^p, q_{Y'}^p \leq q_m + q_a + p$. Since \mathcal{V}'^p and \mathcal{Y}'^p are defined by independent permutations, we know that \mathcal{V}'^p and \mathcal{Y}'^p are disjoint. We also know that

$$V_0 = \{\pi_1(u) \mid (u, v) \in \tau_1\}, \qquad Y_0 = \{\pi_2(x) \mid (x, y) \in \tau_2\}.$$

are already well defined by the system. Hence the only unknowns that are left are in the sets $\mathcal{V}'^p \setminus V_0$ and $\mathcal{Y}'^p \setminus Y_0$. For $v_s \in \mathcal{V}'^p$ and $y_t \in \mathcal{Y}'^p$, we connect v_s and y_t with a λ -labeled edge of label $Z^{(i)}$ if $\pi_1(A^{(i)}) = v_s$ and $\pi_2(B^{(i)}) = y_t$ for some $i \in [q_m]$. Similarly, we connect v_s and y_t with a λ' -labeled edge of label $Z'^{(a)}$ if $\pi_1(A'^{(a)}) = v_s$ and $\pi_2(B'^{(a)}) = y_t$ for some $a \in [q_a]$. Finally, v_s (resp., y_t) represents an isolated colliding vertex if it is not connected with an edge, for these vertices we have $\pi_1(u) = v_s$ (resp., $\pi_2(v) = y_t$) for $(u, v_s) \in \tau_1$ and $(x, y_t) \in \tau_2$. In this way, we obtain a graph on $\mathcal{V}'^p \sqcup \mathcal{Y}'^p$, called the transcript graph of τ , and we denote it by $\mathcal{G}_{\tau}(\mathcal{E}^p)$.

4.3 Bad Transcripts.

Informally, bad events are the properties which would make the public permutation extended mirror theory inapplicable. One can only apply the mirror theory if $\mathcal{G}_{\tau}(\mathcal{E}^p)$ is (1). acyclic, (2). satisfies the non-degeneracy condition, (3). satisfies the NCL condition, (4). satisfies the SCV condition, (5). satisfies the NDL condition, and (6). is $(\xi + 1)$ -block-maximal. For some parameter ξ that will be defined later on, we say a system of equations is $(\xi + 1)$ -block-maximal if it does not contain a $(\xi + 1)$ -block-collision, which means that neither of the two permutations evaluates the same input more than ξ times. As our security analysis will cap on 2n/3-bit security only, we can keep it simple by introducing an event that excludes all alternating paths of length 3, and events that exclude all v- \star component with a y-colliding vertex and y- \star component with a v-colliding vertex. Below, we will give a formal description of the bad events.

For simplicity, we denote by $A^{(i)}$ the *i*-th input to π_1 , $B^{(i)}$ the *i*-th input to π_2 , and $Z^{(i)} = \pi_1(A^{(i)}) \oplus \pi_2(B^{(i)})$ for the MAC queries. Similarly we have $A'^{(a)}$, $B'^{(a)}$ and $Z'^{(a)}$ for the verification queries. Given a parameter $\xi \in \mathbb{N}$, we say that $\tau \in \mathcal{T}_{\text{bad}}$ if and only if one of the following conditions holds:

(i) A component with two colliding vertices.

$$\exists i \in [q_m], (u, v) \in \tau_1, (x, y) \in \tau_2 \text{ such that } A^{(i)} = u \land B^{(i)} = x , \\ \exists i \in [q_m], (u, v) \in \tau_1, (x, y) \in \tau_2 \text{ such that } A^{(i)} = u \land Z^{(i)} = v \oplus y , \\ \exists i \in [q_m], (u, v) \in \tau_1, (x, y) \in \tau_2 \text{ such that } Z^{(i)} = v \oplus y \land B^{(i)} = x .$$

(ii) An alternating path of length 3.

$$\exists i \neq k, k \neq l \in [q_m]$$
 such that $A^{(i)} = A^{(k)} \wedge B^{(k)} = B^{(l)}$.

(iii) An alternating path of length 2 such that $\lambda(\mathcal{L}) = 0$.

$$\exists i \neq k \in [q_m] \text{ such that } A^{(i)} = A^{(k)} \wedge Z^{(i)} = Z^{(k)} ,$$

$$\exists i \neq k \in [q_m] \text{ such that } Z^{(i)} = Z^{(k)} \wedge B^{(i)} = B^{(k)} .$$

(iv) A v- \star colliding component with y-colliding vertices, or a y- \star colliding component with v-colliding vertices.

$$\exists i \neq k \in [q_m], (u, v) \in \tau_1 \text{ such that } A^{(i)} = u \land B^{(i)} = B^{(k)},$$

$$\exists i \neq k \in [q_m], (x, y) \in \tau_2 \text{ such that } B^{(i)} = x \land A^{(i)} = A^{(k)},$$

$$\exists i \neq k \in [q_m], (u, v), (u', v') \in \tau_1 \text{ such that}$$

$$A^{(i)} = u \land A^{(k)} = u' \land v \oplus Z^{(i)} = v' \oplus Z^{(k)},$$

$$\exists i \neq k \in [q_m], (x, y), (x', y') \in \tau_2 \text{ such that}$$

$$B^{(i)} = x \land B^{(k)} = x' \land y \oplus Z^{(i)} = y' \oplus Z^{(k)}.$$

(v) A $(\xi + 1)$ -block-collision.

$$\exists i_1, \dots, i_{\xi+1} \in \{1, \dots, q_m\} \text{ such that } A^{(1)} = \dots = A^{(\xi+1)}, \\ \exists i_1, \dots, i_{\xi+1} \in \{1, \dots, q_m\} \text{ such that } B^{(1)} = \dots = B^{(\xi+1)}.$$

(vi) An alternating circle of length 2 with a λ' -labeled edge.

 $\exists i \in [q_m], a \in [q_a]$ such that $A^{(i)} = A'^{(a)} \wedge B^{(i)} = B'^{(a)} \wedge Z^{(i)} = Z'^{(a)}$.

(vii) A λ' -labeled edge between two vertices with distance λ' .

$$\exists a \in [q_a], (u, v) \in \tau_1, (x, y) \in \tau_2 \text{ such that} A'^{(a)} = u \land B'^{(a)} = x \land Z'^{(a)} = v \oplus y , \exists i \in [q_m], a \in [q_a], (u, v), (u'v') \in \tau_1 \text{ such that} A^{(i)} = u \land B^{(i)} = B'^{(a)} \land A'^{(a)} = u' \land Z'^{(a)} = v \oplus Z^{(i)} \oplus v' , \\ \exists i \in [q_m], a \in [q_a], (x, y), (x', y') \in \tau_2 \text{ such that} B^{(i)} = x \land A^{(i)} = A'^{(a)} \land B'^{(a)} = x' \land Z'^{(a)} = y \oplus Z^{(i)} \oplus y' , \\ \exists i \neq k \in [q_m], a \in [q_a], (u, v) \in \tau_1, (x, y) \in \tau_2 \text{ such that } A^{(i)} = u \\ \land B^{(k)} = x \land B^{(i)} = B'^{(a)} \land A^{(k)} = A'^{(a)} \land Z'^{(a)} = v \oplus Z^{(i)} \oplus y \oplus Z^{(k)}$$

Note that by (ii) and (iv), we will end up with a graph that contains only isolated and v- \star colliding components with a v-colliding vertex, isolated and y- \star colliding components with a y-colliding vertex, v- \star components, y- \star components, isolated components, and isolated vertices. The resulting graph is good since it

- 1. satisfies the SCV condition by conditions (i), (ii), and (iv),
- 2. acyclic by conditions (ii),
- 3. satisfies the non-degeneracy condition by conditions (ii) and (iii),
- 4. is $(\xi + 1)$ -block-maximal by conditions (ii) and (v),
- 5. satisfies the NCL condition by conditions (ii) and (vi),
- 6. satisfies the NDL condition by conditions (ii), (iv), and (vii).

The probability that $\tau \in \mathcal{T}_{bad}$ happens, is given by the sum of the probabilities that each of the above mentioned bad events happens. When the above mentioned events can be excluded in the transcript, then \mathcal{G}_{τ} forms a good transcript graph for $\tau \in \mathcal{T}_{good}$.

4.4 Ratio for Good Transcripts.

Once bad transcripts have been defined, we will show that

$$\Pr[X_{\mathcal{P}} \in \mathcal{T}_{\text{bad}}] \le \varepsilon_{\text{bad}} \,,$$

for a small $\varepsilon_{\text{bad}} > 0$. Next, we fix a good transcript τ . According to (2), we only have to consider $\rho(\tau)/\Pr[\mathcal{P}_m: \mathcal{P}_m \vdash \tau_m \land \mathcal{P}_a \vdash \tau_a]$, with

$$\rho(\tau) = \Pr\left[\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n) \colon \mathcal{O}_m \vdash \tau_m \land \mathcal{O}_a \vdash \tau_a \mid \pi_1 \vdash \tau_1 \land \pi_2 \vdash \tau_2\right].$$

This is exactly the ratio given by Theorem 1 and 2. From (2), we obtain

$$\frac{\Pr[X_{\mathcal{O}} = \tau]}{\Pr[X_{\mathcal{P}} = \tau]} \ge 1 - \varepsilon_{\text{ratio}},$$

and by Lemma 3, we have

$$\mathbf{Adv}_{\mathrm{MAC}}^{\mathrm{mac}}(\mathcal{D}) \leq oldsymbol{E}[arepsilon_{\mathrm{ratio}}] + arepsilon_{\mathrm{bad}}$$
 ,

5 Multi-User Security of Tweakable Even-Mansour Cipher

In this section we consider the 2-round Tweakable Even-Mansour (TEM) construction that was proposed by Cogliati et al. [15]. They showed that 2-round TEM achieves 2n/3-bit security in the single-user setting. Here we show that same level of security can be achieved in the *multi-user* setting using the technique proposed in this work.

Let $n \in \mathbb{N}$, let $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$, and let \mathcal{H} be an ϵ -AXU function family. One can consider a generic construction TEM: $\mathcal{H}^2 \times \mathcal{T} \times \{0,1\}^n \to \{0,1\}^n$ as

$$\text{TEM}(h_1, h_2, T, M) = \pi_2(\pi_1(M \oplus h_1(T)) \oplus h_1(T) \oplus h_2(T)) \oplus h_2(T).$$
(4)

The security of TEM based on π_1 and π_2 is given in the following Theorem.

Theorem 3. Let $n \in \mathbb{N}$ and let \mathcal{H} be a uniform ϵ -AXU family of functions from \mathcal{T} to $\{0,1\}^n$. We consider TEM: $\mathcal{H}^2 \times \mathcal{T} \times \{0,1\}^n \to \{0,1\}^n$ based on two permutations $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$ and u pairs of uniform user hash keys $(h_1^1, h_2^1), \ldots, (h_1^u, h_2^u) \stackrel{\$}{\leftarrow} \mathcal{H}^2$. For any distinguisher \mathcal{D} making at most q construction queries distributed over its u construction oracles, at most p primitive queries to π_1^{\pm} and p primitive queries to π_2^{\pm} , we have

$$\mathbf{Adv}_{\mathrm{TEM}}^{\mathrm{mu-tsprp}}(\mathcal{D}) \leq 3q^{3}\epsilon^{2} + q^{2}p\epsilon^{2} + 6\sqrt{q}p\epsilon + \frac{6q^{3/2}}{2^{n}} + \frac{2q(p+q)^{2}}{2^{2n}}\left(1 + 13q\epsilon\right).$$

Proof. Let $(h_1^1, h_2^1), \ldots, (h_1^u, h_2^u) \stackrel{\$}{\leftarrow} \mathcal{H}^2$, $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$, and $\tilde{\pi}_1, \ldots, \tilde{\pi}_u \stackrel{\$}{\leftarrow} \operatorname{Perm}(t, n)$. Here, we consider any distinguisher \mathcal{D} that has access to either $(\operatorname{TEM}_{h_1^1, h_2^1}^{\pi_1, \pi_2^{-1}}, \ldots, \operatorname{TEM}_{h_1^u, h_2^u}^{\pi_1, \pi_2^{-1}}, \pi_1, \pi_2)$ in the real world, or $(\tilde{\pi}_1, \ldots, \tilde{\pi}_u, \pi_1, \pi_2)$ in the ideal world. The security proof relies on Theorem 1 (a), although this application is not straightforward. Most importantly, we consider $(\operatorname{TEM}_{h_1^j, h_2^j}^{\pi_1, \pi_2^{-1}})_{j=1}^u$ instead of $(\operatorname{TEM}_{h_1^j, h_2^j}^{\pi_1, \pi_2})_{j=1}^u$. As π_1, π_2 are drawn independently, these two constructions are provably equally secure. We can view an evaluation $C = \operatorname{TEM}_{h_1^j, h_2^j}^{\pi_1, \pi_2^{-1}}(T, M)$ as the xor of two public permutations in the middle of the function, $\pi_1(M \oplus h_1^j(T)) \oplus \pi_2(M \oplus h_2^j(T)) = h_1^j(T) \oplus h_2^j(T)$. Therefore, q evaluations of $\operatorname{TEM}_{h_1^j, h_2^j}^{\pi_1, \pi_2^{-1}}$ can be translated to a system of q bi-variate affine equations. Including 2p univariate affine equations that are defined by the primitive queries, these equations can be written in the form (3).

Pr[$X_{\mathcal{P}} \in \mathcal{T}_{bad}$]. Following the framework given in Section 4, we first perform the bad transcripts analysis. By replacing $A = M \oplus h_1^j(T)$, $B = M \oplus h_2^j(T)$, and $Z = h_1^j(T) \oplus h_2^j(T)$ in the framework of Section 4.3, we get the following bad events. Given a parameter $\xi \in \mathbb{N}$, we say that $\tau \in \mathcal{T}_{bad}$ if and only if there exist construction queries $(j, T, M, C), (j', T', M', C'), (j'', T'', M'', C'') \in$ τ_m , primitive queries $(u, v), (u', v') \in \tau_1$ and $(x, y), (x', y') \in \tau_2$ such that one of the following conditions holds: (i) A component with two colliding vertices.

$$\begin{aligned} &\text{bad}_1 \colon M \oplus u = h_1^j(T) \land C \oplus x = h_2^j(T) \,, \\ &\text{bad}_2 \colon M \oplus u = h_1^j(T) \land v \oplus y = h_1^j(T) \oplus h_2^j(T) \,, \\ &\text{bad}_3 \colon v \oplus y = h_1^j(T) \oplus h_2^j(T) \land C \oplus x = h_2^j(T) \,. \end{aligned}$$

(ii) An alternating path of length 3.

bad₄:
$$M \oplus h_1^j(T) = M' \oplus h_1^{j'}(T') \land C' \oplus h_2^{j'}(T') = C'' \oplus h_2^{j''}(T'')$$
.

(iii) An alternating path of length 2 such that $\lambda(\mathcal{L}) = 0$.

bad₅:
$$M \oplus h_1^j(T) = M' \oplus h_1^{j'}(T') \wedge h_1^j(T) \oplus h_2^j(T) = h_1^{j'}(T') \oplus h_2^{j'}(T')$$
,
bad₆: $h_1^j(T) \oplus h_2^j(T) = h_1^{j'}(T') \oplus h_2^{j'}(T') \wedge C \oplus h_2^j(T) = C' \oplus h_2^{j'}(T')$.

(iv) A v- \star colliding component with y-colliding vertices, or a y- \star colliding component with v-colliding vertices.

$$\begin{aligned} \operatorname{bad}_{7} &: M \oplus u = h_{1}^{j}(T) \wedge C \oplus h_{2}^{j}(T) = C' \oplus h_{2}^{j'}(T) \,, \\ \operatorname{bad}_{8} &: C \oplus x = h_{2}^{j}(T) \wedge M \oplus h_{1}^{j}(T) = M' \oplus h_{1}^{j'}(T') \,, \\ \operatorname{bad}_{9} &: M \oplus u = h_{1}^{j}(T) \wedge M' \oplus u' = h_{1}^{j'}(T') \\ & \wedge v \oplus h_{1}^{j}(T) \oplus h_{2}^{j}(T) = v' \oplus h_{1}^{j'}(T') \oplus h_{2}^{j'}(T') \,, \\ \operatorname{bad}_{10} &: C \oplus x = h_{2}^{j}(T) \wedge C' \oplus x' = h_{2}^{j'}(T) \\ & \wedge y \oplus h_{1}^{j}(T) \oplus h_{2}^{j}(T) = y' \oplus h_{1}^{j'}(T') \oplus h_{2}^{j'}(T') \,. \end{aligned}$$

(v) A $(\xi + 1)$ -block-collision.

 $\begin{aligned} \operatorname{bad}_{11} \colon \{i_1, \dots, i_{\xi+1}\} &\in [q] \text{ such that } M_{i_1} \oplus h_1^{j_{i_1}}(T_{i_1}) = \dots = M_{i_{\xi+1}} \oplus h_1^{j_{\xi+1}}(T_{i_{\xi+1}}) \,, \\ \operatorname{bad}_{12} \colon \{i_1, \dots, i_{\xi+1}\} &\in [q] \text{ such that } C_{i_1} \oplus h_2^{j_{i_1}}(T_{i_1}) = \dots = C_{i_{\xi+1}} \oplus h_2^{j_{\xi+1}}(T_{i_{\xi+1}}) \,. \end{aligned}$

Since there is a $\sum_{i=1}^{c_1+c_2} q/2^n$ term in Theorem 1 (a), and we want to get 2n/3-bits security, we also need the following two bad events

$$\begin{aligned} \operatorname{bad}_{c_1} &: c_1 = |(j, T, M, C) \in \tau_m \colon M \oplus h_1^j(T) \in \tau_1| \ge \sqrt{q} \,, \\ \operatorname{bad}_{c_2} &: c_2 = |(j, T, M, C) \in \tau_m \colon C \oplus h_2^j(T) \in \tau_2| \ge \sqrt{q} \,. \end{aligned}$$

Lemma 4. For any integers q and p, one has

$$\Pr[\tau \in \mathcal{T}_{\text{bad}}] \le 4qp^2\epsilon^2 + 3q^3\epsilon^2 + q^2p\epsilon^2 + \frac{q^3}{2^{2n}} + 2\sqrt{q}p\epsilon + \frac{16q^2(p+q)^2\epsilon}{2^{2n}}$$

The proof of the lemma is given in Supplementary Material C.

 $\Pr[X_{\mathcal{O}} = \tau] / \Pr[X_{\mathcal{P}} = \tau]$. The next step is the calculate the ratio for good transcripts. Note that by $\neg \operatorname{bad}_{c_1}$ and $\neg \operatorname{bad}_{c_2}$, we have $\sum_{i=1}^{c_1+c_2} q \leq 2q^{3/2}$. We use Theorem 1 (a) to get

$$\epsilon_{\text{ratio}} \le \frac{2q^{3/2}}{2^n} + \frac{3q^3}{2^{2n}} + \frac{2(p+q)^2 \sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} (\eta_{i+1}^2 + \eta_{i+1})}{2^{2n}} + \frac{2q(p+q)^2}{2^{2n}}$$

Let \sim_1 (resp., \sim_2) be an equivalence relation on [q] as $\alpha \sim_1 \beta$ (resp., $\alpha \sim_2 \beta$) if and only if $A_{\alpha} = A_{\beta}$ (resp. $B_{\alpha} = B_{\beta}$). Now, each η_i random variable denotes the cardinality of some non-singleton equivalence class of [q] with respect to either \sim_1 or \sim_2 . For $r, s \in \mathbb{N}$, we denote by $\mathcal{P}_1^1, \ldots, \mathcal{P}_r^1$ and $\mathcal{P}_1^2, \ldots, \mathcal{P}_s^2$ the non-singleton equivalence classes of [q] with respect to \sim_1 and \sim_2 , respectively. Further, for $k \in [r]$ and $l \in [s]$, let $\nu_k = |\mathcal{P}_k^1|$ and $\nu'_l = |\mathcal{P}_l^2|$. Then, we have

$$E\left[\sum_{i=c_{1}+c_{2}}^{c_{1}+c_{2}+c_{3}+c_{4}-1}(\eta_{i+1}^{2}+\eta_{i+1})\right] \leq E\left[\sum_{k=1}^{r}\nu_{k}^{2}+\nu_{k}\right] + E\left[\sum_{l=1}^{s}\nu_{l}^{\prime 2}+\nu_{l}^{\prime}\right]$$
$$\leq 5q^{2}\epsilon,$$

using Lemma 1 and the fact that $(h_1^1, h_2^1), \ldots, (h_1^u, h_2^u) \stackrel{*}{\leftarrow} \mathcal{H}^2$.

Finally, Theorem 3 is proven by combining Lemma 4 and ϵ_{ratio} with Lemma 3.

6 Multi-User Security of pEDM PRF

In this section we consider the permutation based version of Encrypted Davies-Mayer (pEDM) construction, that was proposed by Dutta et al. [22]. They showed that pEDM based on a single permutation achieves 2n/3-bit security. Here we will prove the *multi-user* security of pEDM based on two independent permutations, and we show that same level of security can be achieved using the technique proposed in this work. In this case, the multi-user security analysis is more complex than the single-user analysis, since the inputs to π_1 do not need to be fresh, this leads to more bad events and a more complex good transcripts ratio analysis when a dedicated proof need to be performed.

Let $n \in \mathbb{N}$, let $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$. One can consider a generic construction pEDM: $\{0,1\}^{2n} \times \{0,1\}^n \to \{0,1\}^n$ as

$$pEDM(K_1, K_2, M) = \pi_2(\pi_1(M \oplus K_1) \oplus M \oplus K_1 \oplus K_2) \oplus K_1.$$
(5)

The security of pEDM based on π_1 and π_2 is given in the following Theorem.

Theorem 4. Let $n \in \mathbb{N}$ and $1 \leq \xi \leq 2^{n-1}/(p+q)$. We consider pEDM: $\{0,1\}^{2n} \times \{0,1\}^n \to \{0,1\}^n$ based on two permutations $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$, and u pairs of uniform user keys $(K_1^1, K_2^1), \ldots, (K_1^u, K_2^u) \stackrel{\$}{\leftarrow} \{0,1\}^{2n}$. For any distinguisher \mathcal{D} making at most q construction queries distributed over its u construction oracles,

at most p primitive queries to π_1^{\pm} and p primitive queries to π_2^{\pm} , we have

$$\begin{aligned} \mathbf{Adv}_{\text{pEDM}}^{\text{mu-prf}}(\mathcal{D}) &\leq \frac{2}{2^n} + \frac{4qp^2}{2^{2n}} + \frac{3q^2p}{2^{2n}} + \frac{3p\sqrt{nq}}{2^n} + \frac{2q^3}{2^{2n}} \\ &+ \frac{p^{3/2}}{2^n} + \frac{(p+q)^2}{2^{2n}} \Big(7q + \frac{5u(u-1)}{2^n}\Big) + \frac{\binom{q}{\xi+1}}{2^{n\xi}}. \end{aligned}$$

Proof. Let $(K_1^1, K_2^1), \ldots, (K_1^u, K_2^u) \stackrel{\$}{\leftarrow} \{0, 1\}^{2n}, \pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$, and $\varphi_1, \ldots, \varphi_u \stackrel{\$}{\leftarrow} \operatorname{Func}(n)$. Here, we consider any distinguisher \mathcal{D} that has access to either $(\operatorname{pEDM}_{K_1^1, K_2^1}^{\pi_1, \pi_2^{-1}}, \ldots, \operatorname{pEDM}_{K_1^u, K_2^u}^{\pi_1, \pi_2}, \pi_1, \pi_2)$ in the real world, or $(\varphi_1, \ldots, \varphi_u, \pi_1, \pi_2)$ in the ideal world. The security proof relies on Theorem 1 (b). As before, we consider $(\operatorname{pEDM}_{K_1^1, K_2^1}^{\pi_1, \pi_2^{-1}})_{j=1}^u$ instead of $(\operatorname{pEDM}_{K_1^j, K_2^j}^{\pi_1, \pi_2})_{j=1}^u$. We can view an evaluation $C = \operatorname{pEDM}_{K_1^j, K_2^j}^{\pi_1, \pi_2^{-1}}(M)$ as the xor of two public permutations in the middle of the function, $\pi_1(M \oplus K_1^j) \oplus \pi_2(C \oplus K_1^j) = M \oplus K_1^j \oplus K_2^j$. Therefore, q evaluations of $\operatorname{pEDM}_{K_1^j, K_2^j}^{\pi_1, \pi_2^{-1}}$ can be translated to a system of q bi-variate affine equations. Including 2p uni-variate affine equations that are defined by the primitive queries.

 $\Pr[X_{\mathcal{P}} \in \mathcal{T}_{bad}]$. Following the framework given in Section 4, we first perform the bad transcripts analysis. By replacing $A = M \oplus K_1^j$, $B = C \oplus K_1^j$, and $Z = M \oplus K_1^j \oplus K_2^j$ in the framework of Section 4.3, we get the following bad events. Given a parameter $\xi \in \mathbb{N}$, we say that $\tau \in \mathcal{T}_{bad}$ if and only if there exist construction queries $(j, M, C), (j', M', C'), (j'', M'', C'') \in \tau_m$, and primitive queries $(u, v), (u', v') \in \tau_1$ and $(x, y), (x', y') \in \tau_2$ such that one of the following conditions holds:

(i) A component with two colliding vertices.

$$\begin{aligned} \mathrm{bad}_1 \colon M \oplus u &= K_1^j \wedge C \oplus x = K_1^j, \\ \mathrm{bad}_2 \colon M \oplus u &= K_1^j \wedge v \oplus y = M \oplus K_1^j \oplus K_2^j, \\ \mathrm{bad}_3 \colon v \oplus y &= M \oplus K_1^j \oplus K_2^j \wedge C \oplus x = K_1^j. \end{aligned}$$

(ii) Alternating paths of length 3 across different users.

$$\operatorname{bad}_4 \colon M \oplus K_1^j = M' \oplus K_1^{j'} \wedge C' \oplus K_1^{j'} = C'' \oplus K_1^{j''}.$$

(iii) Alternating paths of length 2 such that $\lambda(\mathcal{L}) = 0$ across different users.

$$\begin{aligned} \operatorname{bad}_5 &: M \oplus K_1^j = M' \oplus K_1^{j'} \wedge M \oplus K_1^j \oplus K_2^j = M' \oplus K_1^{j'} \oplus K_2^{j'} ,\\ \operatorname{bad}_6 &: M \oplus K_1^j \oplus K_2^j = M' \oplus K_1^{j'} \oplus K_2^{j'} \wedge C \oplus K_1^j = C' \oplus K_1^{j'} . \end{aligned}$$

(iv) A v- \star colliding component with y-colliding vertices, or a y- \star colliding component with v-colliding vertices.

$$\begin{aligned} \mathrm{bad}_7 \colon M \oplus u &= K_1^j \wedge C \oplus K_1^j = C' \oplus K_1^{j'}, \\ \mathrm{bad}_8 \colon C \oplus x = K_1^j \wedge M \oplus K_1^j = M' \oplus K_1^{j'}, \\ \mathrm{bad}_9 \colon M \oplus u &= K_1^j \wedge M' \oplus u' = K_1^{j'} \\ & \wedge v \oplus M \oplus K_1^j \oplus K_2^j = v' \oplus M' \oplus K_1^{j'} \oplus K_2^{j'}, \\ \mathrm{bad}_{10} \colon C \oplus x = K_1^j \wedge C' \oplus x' = K_1^{j'} \\ & \wedge y \oplus M \oplus K_1^j \oplus K_2^{j'} = y' \oplus M' \oplus K_1^j \oplus K_2^{j'}. \end{aligned}$$

(v) A $(\xi + 1)$ -block-collision.

bad₁₁:
$$\{i_1, \ldots, i_{\xi+1}\} \in [q]$$
 such that $C_{i_1} \oplus K_1^{j_{i_1}} = \cdots = C_{i_{\xi+1}} \oplus K_1^{j_{\xi+1}}$.

Note that the events bad_4 - bad_6 and bad_8 will not appear when the *single user* setting is considered, since in that case the distinguisher is not allow to query the same M to the construction oracle.

Lemma 5. Let $1 \le \xi \le 2^{n-1}/(p+q)$. For any integers q and p, one has

$$\Pr[\tau \in \mathcal{T}_{\text{bad}}] \le \frac{2}{2^n} + \frac{4qp^2}{2^{2n}} + \frac{3q^2p}{2^{2n}} + \frac{3p\sqrt{nq}}{2^n} + \frac{2q^3}{2^{2n}} + \frac{p^{3/2}}{2^n} + \frac{\binom{q}{\xi+1}}{2^{n\xi}}$$

The proof of the lemma is given in Supplementary Material D.

 $\Pr[X_{\mathcal{O}} = \tau] / \Pr[X_{\mathcal{P}} = \tau]$. The next step is the calculate the ratio for good transcripts. We use Theorem 1 (b) to get

$$\epsilon_{\text{ratio}} \leq \frac{2(p+q)^2 \sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} (\eta_{i+1}^2+\eta_{i+1})}{2^{2n}} + \frac{2q(p+q)^2}{2^{2n}} \,.$$

As before, for $r, s \in \mathbb{N}$, we denote by $\mathcal{P}_1^1, \ldots, \mathcal{P}_r^1$ and $\mathcal{P}_1^2, \ldots, \mathcal{P}_s^2$ the non-singleton equivalence classes of [q] with respect to \sim_1 and \sim_2 , respectively. Further, for $k \in [r]$ and $l \in [s]$, let $\nu_k = |\mathcal{P}_k^1|$ and $\nu'_l = |\mathcal{P}_l^2|$. Then, we have

$$\begin{split} \mathbf{E}\left[\sum_{i=c_{1}+c_{2}}^{c_{1}+c_{2}+c_{3}+c_{4}-1}(\eta_{i+1}^{2}+\eta_{i+1})\right] &\leq \mathbf{E}\left[\sum_{k=1}^{r}\nu_{k}^{2}+\nu_{k}\right] + \mathbf{E}\left[\sum_{l=1}^{s}\nu_{l}^{\prime 2}+\nu_{l}^{\prime}\right] \\ &\leq \frac{5u(u-1)}{2^{n-1}} + \frac{5q^{2}}{2^{n-1}}\,. \end{split}$$

The non-freshness of π_1 in the multi-user setting leads to the existence of v- \star components ($c_3 \neq 0$). The difficulty introduced by this can easily be handled by our new technique without performing a long and complicated analysis. Note that when u = 1, we are back to the single user setting, then there are no v- \star components ($c_3 = 0$), since M is always different. Finally, Theorem 4 is proven by combining Lemma 5 and ϵ_{ratio} with Lemma 3.

7 Multi-User Security of $nEHtM_p$ MAC Algorithm

In this section we consider the public permutation based nonce-based Enhance Hash-then-Mask (nEHtM_p) MAC algorithm, that was proposed by Dutta and Nandi [20]. They showed that nEHtM_p based on a single permutation (using domain separation) achieves 2n/3-bit security when the number of faulty nonces μ is sufficiently smaller than $2^{n/3}$. However, according to the framework given in Section 4.3, the authors missed some bad events in their analysis, namely the two last events of (iv) and the three last events of (vii) of Section 4.3. Taking into account these missing bad events, the extended mirror theory used in [20] is *not* sufficient for the good transcripts ratio analysis of nEHtM_p. As a result, their ratio analysis of the construction is also incomplete. More precisely, nonequations between a colliding component and a normal component were not considered in the ratio analysis of [20]. This observation was also verified by the authors [10].

In this section, we will fix this problem using the techniques proposed in this work without performing a new complicated analysis, since these non-equations are already covered in our public permutation extended mirror theory (Theorem 2). Some of these additional bad events, however, require involved arguments to bound. Since the goal of this work is to illustrate the power of our new modular proof approaches, rather than presenting strong combinatorial results to bound these events. We will modify the design of $nEHtM_p$ by xoring an universal hash evaluation of the input message M using an extra hash key h^* to the output tag. This modified $m-nEHtM_p$ construction uses more randomness, which in turn enables us to bound the additional bad events easily. We would like to note that our analysis of $m-nEHtM_p$ does not imply infeasibility in fixing the proof of $nEHtM_p$. In fact, we believe that the security of the original $nEHtM_p$ construction can also be proven with our new approaches in combination with some strong techniques to bound these bad events. Here we will prove that this m-nEHtM_p construction based on two independent permutations achieves 2n/3bit security in the *multi-user* setting using the technique proposed in this work.

Let $n \in \mathbb{N}$, let $\pi_1, \pi_2 \stackrel{\hspace{0.1em}\mathsf{\leftarrow}}{\leftarrow} \operatorname{Perm}(n)$, and let \mathcal{H} be an ϵ -AXU function family. One can consider a generic construction m-nEHtM_p: $\{0,1\}^n \times \mathcal{H}^2 \times \{0,1\}^n \times \mathcal{M} \to \{0,1\}^n$ as

m-nEHtM_p(K, h, h*N, M) =
$$\pi_1(N \oplus K) \oplus \pi_2(N \oplus h(M)) \oplus h^*(M)$$
. (6)

The security of m-nEHtM_p based on π_1 and π_2 is given in the following Theorem.

Theorem 5. Let $n \in \mathbb{N}$, and let \mathcal{H} be a uniform ϵ -AXU family of functions from \mathcal{M} to $\{0,1\}^n$. We consider m-nEHtM_p: $\{0,1\}^n \times \mathcal{H}^2 \times \{0,1\}^n \times \mathcal{M} \to \{0,1\}^n$ based on two permutations $\pi_1, \pi_2 \stackrel{\$}{\leftarrow} \operatorname{Perm}(n)$, u uniform user keys $K_1, \ldots, K_u \stackrel{\$}{\leftarrow} \{0,1\}^n$ and u pairs of uniform user hash keys $(h_1,h_1^*), \ldots, (h_u,h_u^*) \stackrel{\$}{\leftarrow} \mathcal{H}^2$. Let μ be a fixed parameter. For any distinguisher \mathcal{D} making at most q_m queries with at most μ faulty nonces distributed over its u construction MAC oracles, q_a queries

distributed over its u construction verification oracles, at most p primitive queries to π_1^{\pm} and p primitive queries to π_2^{\pm} , we have

$$\begin{aligned} \mathbf{Adv}_{\mathrm{m-nEHtM}_{p}}^{\mathrm{mu-mac}}(\mathcal{D}) &\leq 7\sqrt{q_{m}}p\epsilon + 2\mu^{2}\epsilon + 4q_{m}^{3}\epsilon^{2} + \frac{q_{m}^{2}p\epsilon}{2^{n}} + 2\mu p\epsilon + q_{m}^{2}q_{a}\epsilon^{2} + \frac{q_{a}p^{2}\epsilon}{2^{n}} \\ &+ \frac{3q_{m}q_{a}p\epsilon}{2^{n}} + p\sqrt{q_{m}q_{a}}\epsilon^{\frac{3}{2}} + \frac{(p+q_{m})^{2}}{2^{2n}} \Big(5\mu^{2} + 7q_{m} + \frac{5u(u-1)}{2^{n}}\Big) + \frac{q_{a}}{2^{n}}. \end{aligned}$$

Proof. Let $K_1, \ldots, K_u \stackrel{\$}{\leftarrow} \{0,1\}^n$, $(h_1, h_1^*), \ldots, (h_u, h_u^*) \stackrel{\$}{\leftarrow} \mathcal{H}^2$, and $\pi_1, \pi_2 \stackrel{\$}{\leftarrow}$ Perm(n). Here, we consider any distinguisher \mathcal{D} that has access to either $(\mathcal{O}, \pi_1, \pi_2)$ in the real world or $(\mathcal{P}, \pi_1, \pi_2)$ in the ideal world, with $\mathcal{O} = \left((\text{m-nEHtM}_{p(K_1, h_1, h_1^*)}^{\pi_1, \pi_2}, \text{Ver}_{(K_1, h_1, h_1^*)}^{\pi_1, \pi_2}) \right)$ and $\mathcal{P} = \left((\text{Rand}_1, \text{Rej}_1), \ldots, (\text{Rend}_u, \text{Rej}_u) \right)$. The security proof relies on Theorem 2 (b).

Pr[$X_{\mathcal{P}} \in \mathcal{T}_{bad}$]. Following the framework given in Section 4, we first perform the bad transcripts analysis. For the notational simplicity, we denote $H_j = h_j(M)$, and By replacing $A = N \oplus K_j$, $B = N \oplus h_j(M)$, and $Z = T \oplus h_j^*(M)$ in Section 4.3, we get the following bad events. Given a parameter $\xi \in \mathbb{N}$, we say that $\tau \in \mathcal{T}_{bad}$ if and only if there exist construction MAC queries $(j, N, M, T), (j', N', M', T'), (j'', N'', M'', T'') \in \tau_m$, a construction verification query $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)}, b^{(a)}) \in \tau_a$ and primitive queries $(u, v), (u', v') \in \tau_1$ and $(x, y), (x', y') \in \tau_2$ such that one of the following conditions holds:

(i) A component with two colliding vertices.

$$bad_1: N \oplus u = K_j \land N \oplus x = H_j,$$

$$bad_2: N \oplus u = K_j \land v \oplus y = T \oplus H_j^*,$$

$$bad_3: v \oplus y = T \oplus H_i^* \land N \oplus x = H_j.$$

(ii) An alternating path of length 3.

bad₄:
$$N \oplus K_j = N' \oplus K_{j'} \land N' \oplus H_{j'} = N'' \oplus H_{j''}$$

(iii) An alternating path of length 2 such that $\lambda(\mathcal{L}) = 0$.

$$bad_5 \colon N \oplus K_j = N' \oplus K_{j'} \wedge T \oplus H_j^* = T' \oplus H_{j'}^*,$$

$$bad_6 \colon T \oplus H_j^* = T' \oplus H_{j'}^* \wedge N \oplus H_{j'} = N' \oplus H_{j'}.$$

(iv) A v- \star colliding component with y-colliding vertices, or a y- \star colliding component with v-colliding vertices.

$$\begin{aligned} \operatorname{bad}_{7} \colon N \oplus u &= K_{j} \wedge N \oplus H_{j} = N' \oplus H_{j'}, \\ \operatorname{bad}_{8} \colon N \oplus x = H_{j} \wedge N \oplus K_{j} = N' \oplus K_{j'}, \\ \operatorname{bad}_{9} \colon N \oplus u &= K_{j} \wedge N' \oplus u' = K_{j'} \wedge v \oplus T \oplus H_{j}^{*} = v' \oplus T' \oplus H_{j'}^{*}, \\ \operatorname{bad}_{10} \colon N \oplus x = H_{j} \wedge N' \oplus x' = H_{j'} \wedge y \oplus T \oplus H_{j}^{*} = y' \oplus T' \oplus H_{j'}^{*}. \end{aligned}$$

(v) A $(\xi + 1)$ -block-collision.

 $\text{bad}_{11}: \{i_1, \dots, i_{\xi+1}\} \in [q_m] \text{ such that } N_{i_1} \oplus H_{j_{i_1}} = \dots = N_{i_{\xi+1}} \oplus H_{j_{i_{\xi+1}}}.$

(vi) An alternating circle of length 2 with a λ' -labeled edge.

bad₁₂:
$$N \oplus K_j = N^{(a)} \oplus K_{j^{(a)}} \wedge N \oplus H_j = N^{(a)} \oplus H_{j^{(a)}}$$

 $\wedge T \oplus H_j^* = T^{(a)} \oplus H_{j^{(a)}}^*$.

(vii) A λ' -labeled edge between two vertices with distance λ' .

 $\begin{aligned} \operatorname{bad}_{13} \colon N^{(a)} \oplus u &= K_{j^{(a)}} \wedge N^{(a)} \oplus x = H_{j^{(a)}} \wedge T^{(a)} \oplus H_{j^{(a)}}^* = v \oplus y \,, \\ \operatorname{bad}_{14} \colon N \oplus u &= K_j \wedge N \oplus H_j = N^{(a)} \oplus H_{j^{(a)}} \wedge N^{(a)} \oplus u' = K_{j^{(a)}} \\ & \wedge T^{(a)} \oplus H_{j^{(a)}}^* = v \oplus T \oplus H_j^* \oplus v' \,, \\ \operatorname{bad}_{15} \colon N \oplus x = H_j \wedge N \oplus K_j = N^{(a)} \oplus K_{j^{(a)}} \wedge N^{(a)} \oplus x' = H_{j^{(a)}} \\ & \wedge T^{(a)} \oplus H_{j^{(a)}}^* = y \oplus T \oplus H_j^* \oplus y' \,, \\ \operatorname{bad}_{16} \colon N \oplus u = K_j \wedge N' \oplus x = H_{j'} \wedge N \oplus H_j = N^{(a)} \oplus H_{j^{(a)}} \wedge \end{aligned}$

$$N' \oplus K_{j'} = N^{(a)} \oplus K_{j^{(a)}} \wedge T^{(a)} \oplus H^*_{j^{(a)}} = v \oplus T \oplus H^*_j \oplus y \oplus T' \oplus H^*_{j'}$$

Note that the events bad_9-bad_{10} and $bad_{14}-bad_{16}$ are the missing events that were not considered by the authors of [20].

Lemma 6. For any integers q_m , q_a and p, then one has

$$\begin{aligned} \Pr[\tau \in \mathcal{T}_{\text{bad}}] &\leq 7\sqrt{q_m}p\epsilon + 2\mu^2\epsilon + 4q_m^3\epsilon^2 + \frac{q_m^2p\epsilon}{2^n} + 2\mu p\epsilon \\ &+ \frac{8q_m^2(p+q_m)^2\epsilon}{2^{2n}} + q_m^2q_a\epsilon^2 + \frac{q_ap^2\epsilon}{2^n} + \frac{3q_mq_ap\epsilon}{2^n} + p\sqrt{q_mq_a}\epsilon^{3/2} \,. \end{aligned}$$

The proof of this Lemma is given in Supplementary Material E.

 $\Pr[X_{\mathcal{O}} = \tau] / \Pr[X_{\mathcal{P}} = \tau]$. The next step is the calculate the ratio for good transcripts. We use Theorem 2 (b) to get

$$\epsilon_{\text{ratio}} \le \frac{2(p+q_m)^2 \sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} (\eta_{i+1}^2+\eta_{i+1})}{2^{2n}} + \frac{2q_m(p+q_m)^2}{2^{2n}} + \frac{2q_a}{2^n}$$

As before, for $r, s \in \mathbb{N}$, we denote by $\mathcal{P}_1^1, \ldots, \mathcal{P}_r^1$ and $\mathcal{P}_1^2, \ldots, \mathcal{P}_s^2$ the non-singleton equivalence classes of $[q_m]$ with respect to \sim_1 and \sim_2 , respectively. For $k \in [r]$ and $l \in [s]$, let $\nu_k = |\mathcal{P}_k^1|$ and $\nu'_l = |\mathcal{P}_l^2|$. Then, we have

$$\begin{split} \mathbf{E} \left[\sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} (\eta_{i+1}^2 + \eta_{i+1}) \right] &\leq \mathbf{E} \left[\sum_{k=1}^r \nu_k^2 + \nu_k \right] + \mathbf{E} \left[\sum_{l=1}^s \nu_l'^2 + \nu_l' \right] \\ &\leq \frac{5\mu^2}{2} + \frac{5u(u-1)}{2^{n-1}} + \frac{5q_m^2\epsilon}{2} \,, \end{split}$$

using Lemma 1 and the fact that $K_1, \ldots, K_u \stackrel{*}{\leftarrow} \{0, 1\}^n$ and $(h_1, h_1^*), \ldots, (h_u, h_u^*) \stackrel{*}{\leftarrow} \mathcal{H}^2$. Note that when u = 1, we are back to the single-user setting, in this case the v- \star components (collision in the inputs of π_1) can only be formed by queries with repeated nonces, hereby the $5\mu^2/2$ term in the bound. Finally, Theorem 5 is proven by combining Lemma 6 and ϵ_{ratio} with Lemma 3.

ACKNOWLEDGMENTS. This work was supported in part by the Research Council KU Leuven: GOA TENSE (C16/15/058). The author was supported by a Ph.D. Fellowship from the Research Foundation - Flanders (FWO). I want to thank Tim Beyne and the reviewers for their valuable comments and suggestions.

References

- NIST Lightweight Cryptography https://csrc.nist.gov/Projects/ Lightweight-Cryptography
- NIST SHA-3 Project https://csrc.nist.gov/projects/hash-functions/ sha-3-project
- Bellare, M., Boldyreva, A., Micali, S.: Public-key encryption in a multi-user setting: Security proofs and improvements. In: EUROCRYPT 2000. LNCS, vol. 1807, pp. 259–274
- Biham, E.: How to decrypt or even substitute des-encrypted messages in 2²⁸ steps. Inf. Process. Lett. 84(3), 117–124
- Bogdanov, A., Knežević, M., Leander, G., Toz, D., Varici, K., Verbauwhede, I.: Spongent: A lightweight hash function. In: CHES 2011. LNCS, vol. 6917, pp. 312– 325
- Bogdanov, A., Knudsen, L.R., Leander, G., Standaert, F.X., Steinberger, J.P., Tischhauser, E.: Key-alternating ciphers in a provable setting: Encryption using a small number of public permutations - (extended abstract). In: EURO-CRYPT 2012. LNCS, vol. 7237, pp. 45–62
- Chatterjee, S., Menezes, A., Sarkar, P.: Another look at tightness. In: SAC 2011. LNCS, vol. 7118, pp. 293–319
- Chen, S., Lampe, R., Lee, J., Seurin, Y., Steinberger, J.P.: Minimizing the tworound Even-Mansour cipher. In: CRYPTO 2014, Part I. LNCS, vol. 8616, pp. 39–56
- Chen, S., Steinberger, J.P.: Tight security bounds for key-alternating ciphers. In: EUROCRYPT 2014. LNCS, vol. 8441, pp. 327–350
- Chen, Y.L., Dutta, A., Nandi, M.: Multi-user BBB security of public permutations based MAC. Cryptogr. Commun. 14(5), 1145–1177
- Chen, Y.L., Lambooij, E., Mennink, B.: How to build pseudorandom functions from public random permutations. In: CRYPTO 2019, Part I. LNCS, vol. 11692, pp. 266–293
- Chen, Y.L., Tessaro, S.: Better security-efficiency trade-offs in permutation-based two-party computation. In: ASIACRYPT 2021, Part II. LNCS, vol. 13091, pp. 275–304
- 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

- 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
- Cogliati, B., Lampe, R., Seurin, Y.: Tweaking Even-Mansour ciphers. In: CRYPTO 2015, Part I. LNCS, vol. 9215, pp. 189–208
- Cogliati, B., Seurin, Y.: EWCDM: An efficient, beyond-birthday secure, noncemisuse resistant MAC. In: CRYPTO 2016, Part I. LNCS, vol. 9814, pp. 121–149
- Daemen, J., Rijmen, V.: The Design of Rijndael The Advanced Encryption Standard (AES), Second Edition. Information Security and Cryptography, Springer (2020)
- Datta, N., Dutta, A., Nandi, M., Yasuda, K.: Encrypt or decrypt? to make a singlekey beyond birthday secure nonce-based MAC. In: CRYPTO 2018, Part I. LNCS, vol. 10991, pp. 631–661
- Dutta, A.: Minimizing the two-round tweakable even-mansour cipher. In: ASI-ACRYPT 2020, Part I. LNCS, vol. 12491, pp. 601–629
- Dutta, A., Nandi, M.: BBB secure nonce based MAC using public permutations. In: AFRICACRYPT 2020. LNCS, vol. 12174, pp. 172–191
- 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
- Dutta, A., Nandi, M., Talnikar, S.: Permutation based EDM: an inverse free BBB secure PRF. IACR Trans. Symmetric Cryptol. 2021(2), 31–70
- Even, S., Mansour, Y.: A construction of a cipher from a single pseudorandom permutation. In: ASIACRYPT'91. LNCS, vol. 739, pp. 210–224
- Guo, C., Katz, J., Wang, X., Yu, Y.: Efficient and secure multiparty computation from fixed-key block ciphers. In: 2020 IEEE Symposium on Security and Privacy. pp. 825–841
- Guo, J., Peyrin, T., Poschmann, A.: The PHOTON family of lightweight hash functions. In: CRYPTO 2011. LNCS, vol. 6841, pp. 222–239
- Hoang, V.T., Tessaro, S.: Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In: CRYPTO 2016, Part I. LNCS, vol. 9814, pp. 3–32
- Jha, A., Nandi, M.: Tight security of cascaded LRW2. Journal of Cryptology 33(3), 1272–1317
- 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
- Krawczyk, H.: LFSR-based hashing and authentication. In: CRYPTO'94. LNCS, vol. 839, pp. 129–139
- Mennink, B., Neves, S.: Encrypted Davies-Meyer and its dual: Towards optimal security using mirror theory. In: CRYPTO 2017, Part III. LNCS, vol. 10403, pp. 556–583
- Mouha, N., Luykx, A.: Multi-key security: The Even-Mansour construction revisited. In: CRYPTO 2015, Part I. LNCS, vol. 9215, pp. 209–223
- Nandi, M.: Mind the composition: Birthday bound attacks on EWCDMD and SoKAC21. In: EUROCRYPT 2020, Part I. LNCS, vol. 12105, pp. 203–220
- Patarin, J.: The "coefficients H" technique (invited talk). In: SAC 2008. LNCS, vol. 5381, pp. 328–345
- Patarin, J.: Mirror theory and cryptography. Appl. Algebra Eng. Commun. Comput. 28(4), 321–338

Supplementary Material

A Proof of Theorem 1

Since the isolated vertices in \mathcal{I} are already well defined, we focus only on the vertices in \mathcal{A} , \mathcal{B} , and \mathcal{C} . For readability, we use $N = 2^n$ and \mathcal{G} instead of $\mathcal{G}(\mathcal{E}_m^p)$. Let $i = 1, \ldots, c$ with $c = c_1 + c_2 + c_3 + c_4 + c_5$. Hence, we assume that \mathcal{A}_1 is the first component and \mathcal{C}_{c_5} is the last component.

- Let ξ_i be the size (number of vertices) of the *i*-th component, and $\eta_i = \xi_i 1$ (number of edges in the *i*-th component).
- for $j \in [\eta_i]$ and $r = \sum_{k=1}^{i-1} \eta_k + j$,
 - $\lambda_i^j = \lambda^r$ (λ value corresponding to the *j*-th equation of *i*-th component).
 - $\delta_i^j := \mu(\lambda^{r-1}, \lambda_i^j)$ with $\delta_1^1 = 0$ for prp-like settings.
 - $\delta_i^j := 0$ for prf-like settings.

When the *i*-th component only contains one edge, we drop the superscript j, and use λ_i and δ_i instead of λ_i^j and δ_i^j .

- For each component, where one part consists of the vertices in \mathcal{V}^p and the other vertices in \mathcal{Y}^p ; we denote the two parts respectively by V_i and Y_i , and we have $\xi_i = |V_i| + |Y_i|$. We denote by v_i^j (resp., y_i^j) the *j*-th vertex in V_i (resp., Y_i). When V_i (resp., Y_i) only contains one vertex, we will drop the superscript, and write v_i (resp., y_i).

Recall there are exactly p colliding v vertices and p colliding y vertices in \mathcal{G} ; and that q_1, q_2, q_3, q_4 , and q_5 are the number of equations (edges) in $\mathcal{A}_1 \sqcup \cdots \sqcup \mathcal{A}_{c_1}$, $\mathcal{A}_{c_1+1} \sqcup \cdots \sqcup \mathcal{A}_{c_1+c_2}, \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_{c_3}, \mathcal{B}_{c_3+1} \sqcup \cdots \sqcup \mathcal{B}_{c_3+c_4}$, and \mathcal{C} , respectively.

Colliding Components. Recall that for a good graph, there is only one colliding vertex in each colliding component (Definition 5). For $i = 1, \ldots, c_1, A_i$ contains one colliding v vertex and η_i unknown y vertices ($\eta_i = 1$ in the case of isolated components). For $i = c_1 + 1, \ldots, c_1 + c_2, A_i$ contains one colliding y vertex and η_i unknown v vertices ($\eta_i = 1$ in the case of isolated components).

For $i = 1, \ldots, c_1 + c_2$, let $h_A(i)$ be the number of solutions to $\mathcal{I} \sqcup \mathcal{A}_1 \sqcup \cdots \sqcup \mathcal{A}_i$. Then we have that $h_A(i) = 1$ for $1 \leq i \leq c_1 + c_2$. For $1 \leq i \leq c_1$, let $v_i^* \in V_i$ be the colliding vertex of \mathcal{A}_i such that the value of this vertex is well defined, then the other η_i unknown y vertices in \mathcal{A}_i are uniquely determined, since each y vertex in \mathcal{A}_i is connected with a λ -labeled edge to v_i^* . Similarly, for $c_1 + 1 \leq i \leq c_1 + c_2$, let $y_i^* \in Y_i$ be the colliding vertex of \mathcal{A}_i such that the value of this vertex is well defined, then the other η_i unknown v vertices in \mathcal{A}_i are uniquely determined, since each y is connected with a λ -labeled edge to v_i^* . Then, we

$$\frac{h_A(c_1+c_2)\prod_{i=1}^{c_1+c_2}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{(N-p)_{q_2}(N-p)_{q_1}} \ge \frac{\prod_{i=1}^{c_1+c_2}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{N^{q_1+q_2}} \\
\ge \frac{N^{\sum_{i=1}^{c_1+c_2}\eta_i} - \sum_{i=1}^{c_1+c_2}\sum_{j=1}^{\eta_i}\delta_i^j N^{\sum_{i=1}^{c_1+c_2}\eta_i-1}}{N^{q_1+q_2}} \\
\ge \frac{N^{\sum_{i=1}^{c_1+c_2}\eta_i}(1-\sum_{i=1}^{c_1+c_2}\sum_{j=1}^{\eta_i}\delta_i^j/N)}{N^{q_1+q_2}} \\
\ge 1-\frac{\sum_{i=1}^{c_1+c_2}\delta}{N},$$
(7)

using $\sum_{i=1}^{c_1+c_2} \eta_i = q_1 + q_2$, and $\sum_{j=1}^{\eta_i} \delta_i^j \leq \delta$. Here we have $\delta = q_m$ in the case of prp-like settings, and $\delta = 0$ in the case of prf-like settings.

For the further analysis, we define

$$\mathcal{V}_{i} = \bigsqcup_{c_{1}+1 \leq i \leq c_{1}+c_{2}} V_{i} \sqcup \bigsqcup_{c_{1}+c_{2}+1 \leq k \leq i} V_{k}$$
$$\mathcal{Y}_{i} = \bigsqcup_{1 \leq i \leq c_{1}} Y_{i} \sqcup \bigsqcup_{c_{1}+c_{2}+1 \leq k \leq i} Y_{k}.$$

Star Components. It can be easily seen that the analysis of the v- \star components and that of the y- \star components are symmetric, so we only perform the analysis for the v-* components. Let $i = 1, \ldots, c_3$ and $i' = c_1 + c_2 + i$, and let $h_B(i)$ be the number of solutions to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_i$. Note that $h_B(0) = h(c_1 + c_2)$. Our goal is to derive a recursive formula for $h_B(i+1)$ that depends on $h_B(i)$, such that a lower bound can be found for the expression $h_B(i+1)/h_B(i)$. In order to do that, we fix a solution to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_i$. If we fix the single vertex $v_{i'+1} \in V_{i'+1}$ and assign any value to $v_{i'+1}$, then the other $\eta_{i'+1}$ unknowns $y_{i'+1}^j$ in the (i'+1)-th component (\mathcal{B}_{i+1}) are uniquely determined, since each $y_{i'+1}^j$ vertex in Y_{i+1} is connected with a λ -labeled edge to $v_{i'+1}$, for $j \in [\eta_{i'+1}]$. Then, $h_B(i+1)$ is the number of solutions

$$\left(V_0 \sqcup \mathcal{V}_{i'} \sqcup v_{i'+1}\right) \sqcup \left(Y_0 \sqcup \mathcal{Y}_{i'} \sqcup \{y_{i'+1}^1, \dots, y_{i'+1}^{\eta_{i'+1}}\}\right)$$

that satisfies:

- (1) $(V_0 \sqcup \mathcal{V}_{i'}) \sqcup (Y_0 \sqcup \mathcal{Y}_{i'})$ satisfies $h_B(i)$; (2) $v_{i'+1} \oplus y_{i'+1}^j = \lambda_{i'+1}^j$ for $j \in [\eta_{i'+1}]$; (3) $v_{i'+1} \notin (V_0 \sqcup \mathcal{V}_{i'})$;

- (4) $y_{i'+1}^{j} \notin (Y_0 \sqcup \mathcal{Y}_{i'}), \text{ for } j \in [\eta_{i'+1}].$

Clearly, there are in total $Nh_B(i)$ solutions that only comply with the conditions (1) and (2), which do not take the conditions (3) and (4) into account. On the other hand, condition (3) excludes $|V_0 \sqcup \mathcal{V}_{i'}|$ possible solutions for $v_{i'+1}$. While condition (4) excludes $|Y_0 \sqcup \mathcal{Y}_{i'}|$ possible solutions for each $y_{i'+1}^j$, hence, (4)

have

excludes at most $|Y_0 \sqcup \mathcal{Y}_{i'}| \eta_{i'+1}$ possible solutions. Further, there are exactly $\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j$ previous equations that share the same λ value with some equation in \mathcal{B}_{i+1} . By the principle of inclusion-exclusion, we obtain

$$h_B(i+1) \ge \left(N - (p + |\mathcal{V}_{i'}|) - (p + |\mathcal{Y}_{i'}|)\eta_{i'+1} + \sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j\right) h_B(i).$$

Then for $0 \le i \le c_3 - 1$ and $i' = c_1 + c_2 + i$, we have

$$\frac{h_B(i+1)\prod_{j=1}^{\eta_{i'+1}}(N-\delta_{i'+1}^j)}{h_B(i)(N-(p+|\mathcal{V}_{i'}|))(N-(p+|\mathcal{Y}_{i'}|))_{\eta_{i'+1}}} \\
\geq \frac{\prod_{j=1}^{\eta_{i'+1}}(N-\delta_{i'+1}^j)\left(N-(p+|\mathcal{V}_{i'}|)-(p+|\mathcal{Y}_{i'}|)\eta_{i'+1}+\sum_{k=1}^{\eta_{i'+1}}\delta_{i'+1}^j\right)}{(N-(p+|\mathcal{V}_{i'}|))(N-(p+|\mathcal{Y}_{i'}|))_{\eta_{i'+1}}} \\
\geq 1-\frac{B-C}{(N-(p+|\mathcal{V}_{i'}|))(N-(p+|\mathcal{Y}_{i'}|))_{\eta_{i'+1}}},$$
(8)

where

$$B = (N - (p + |\mathcal{V}_{i'}|))(N - (p + |\mathcal{Y}_{i'}|))_{\eta_{i'+1}}$$
$$C = \prod_{j=1}^{\eta_{i'+1}} (N - \delta_{i'+1}^j) \left(N - (p + |\mathcal{V}_{i'}|) - (p + |\mathcal{Y}_{i'}|)\eta_{i'+1} + \sum_{k=1}^{\eta_{i'+1}} \delta_{i'+1}^j\right).$$

We need both a lower and an upper bound for *B*. Using the facts that $|\mathcal{V}_{i'}|$, $|\mathcal{Y}_{i'}| + \eta_{i'+1} < q_m$, and $\xi_{\max}(p+q_m) < 2^{n-1}$, we get $B \ge N^{(\eta_{i'+1}+1)}/2$. We now derive an upper bound for *B*.

$$\begin{split} B &\leq (N - (p + |\mathcal{V}_{i'}|))(N - (p + |\mathcal{Y}_{i'}|))^{\eta_{i'+1}} \\ &\leq (N - (p + |\mathcal{V}_{i'}|)) \Big(N^{\eta_{i'+1}} - \eta_{i'+1}(p + |\mathcal{Y}_{i'}|) N^{\eta_{i'+1}-1} + \eta^2_{i'+1}(p + |\mathcal{Y}_{i'}|)^2 N^{\eta_{i'+1}-2} \Big) \\ &\leq N^{\eta_{i'+1}+1} - \eta_{i'+1}(p + |\mathcal{Y}_{i'}|) N^{\eta_{i'+1}} + \eta^2_{i'+1}(p + |\mathcal{Y}_{i'}|)^2 N^{\eta_{i'+1}-1} \\ &- (p + |\mathcal{V}_{i'}|) N^{\eta_{i'+1}} + \eta_{i'+1}(p + |\mathcal{V}_{i'}|)(p + |\mathcal{Y}_{i'}|) N^{\eta_{i'+1}-1} \,. \end{split}$$

We now derive a lower bound for C.

$$C \ge \left(N^{\eta_{i'+1}} - \sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j N^{\eta_{i'+1}-1}\right) \left(N - (p + |\mathcal{V}_{i'}|) - (p + |\mathcal{Y}_{i'}|)\eta_{i'+1} + \sum_{k=1}^{\eta_{i'+1}} \delta_{i'+1}^j\right)$$
$$\ge N^{\eta_{i'+1}+1} - (p + |\mathcal{V}_{i'}|)N^{\eta_{i'+1}} - \eta_{i'+1}(p + |\mathcal{Y}_{i'}|)N^{\eta_{i'+1}} - \left(\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j\right)^2 N^{\eta_{i'+1}-1}.$$

On substituting the bounds of B and C in (8), we get

$$(8) \geq 1 - \frac{\eta_{i'+1}^{2}(p+|\mathcal{Y}_{i'}|)^{2}N^{\eta_{i'+1}-1} + \eta_{i'+1}(p+|\mathcal{V}_{i'}|)(p+|\mathcal{Y}_{i'}|)N^{\eta_{i'+1}-1}}{N^{\eta_{i'+1}+1}/2} - \frac{\left(\sum_{j=1}^{\eta_{i'+1}}\delta_{i'+1}^{j}\right)^{2}N^{\eta_{i'+1}-1}}{N^{\eta_{i'+1}+1}/2} \\\geq 1 - \frac{\eta_{i'+1}^{2}(p+q_{m})^{2}N^{\eta_{i'+1}-1} + \eta_{i'+1}(p+q_{m})^{2}N^{\eta_{i'+1}-1} - \left(\sum_{j=1}^{\eta_{i'+1}}\delta_{i'+1}^{j}\right)^{2}N^{\eta_{i'+1}-1}}{N^{\eta_{i'+1}+1}/2} \\\geq 1 - \frac{2(p+q_{m})^{2}(\eta_{i'+1}^{2} + \eta_{i'+1}) + 2\delta^{2}}{N^{2}},$$

using the fact that $|\mathcal{V}_{i'}| = q_2 + c_3 + q_4 \leq q_m, |\mathcal{Y}_{i'}| = q_1 + q_3 + c_4 \leq q_m$, and $\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j \leq \delta$. Here we have $\delta = q_m$ in the case of prp-like settings $(\lambda_{i'+1}^j \text{ can occur at most once in any component})$, and $\delta = 0$ in the case of prf-like settings.

We consider the $c_3 v \star$ and the $c_4 y \star$ components together

$$\frac{h_B(c_3+c_4)\prod_{i=c_1+c_2+t_3}^{c_1+c_2+c_3+c_4}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{(N-p-q_2)_{c_3+q_4}(N-p-q_1)_{q_3+c_4}} \\
\geq \prod_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} \left(1-\frac{2(p+q_m)^2(\eta_{i+1}^2+\eta_{i+1})+2\delta^2}{N^2}\right) \\
\geq 1-\sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} \left(\frac{2(p+q_m)^2(\eta_{i+1}^2+\eta_{i+1})+2\delta^2}{N^2}\right) \\
\geq 1-\frac{2(p+q_m)^2\sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1}(\eta_{i+1}^2+\eta_{i+1})+q_m\delta^2}{N^2},$$
(9)

using the fact that $c_3 + c_4 \le (q_3 + q_4)/2 \le q_m/2$.

Isolated components. We now consider the components of the graph that contain exactly two vertices. Let $c' = c_1 + c_2 + c_3 + c_4$. For $i = 1, \ldots, c_5$ and $i'' = c_1 + c_2 + c_3 + c_4 + i$, we will write

$$\mathcal{C}_i: v_{i''} \stackrel{\lambda_{i''}}{-} y_{i'}$$

where $v_{i''} \in V_{i''}$ and $y_{i''} \in Y_{i''}$. Let $h_C(i)$ be the number of solutions to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B} \sqcup \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$. Note that $h_C(0) = h(c_1 + c_2 + c_3 + c_4)$ and $h_C(c_5) = h(\mathcal{G})$. Our goal is to derive a recursive formula for $h_C(i+1)$ that depends on $h_C(i)$, such that a lower bound can be found for the expression $h_C(i+1)/h_C(i)$. In order to do that, we fix a solution to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B} \sqcup \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$. Then, $h_C(i+1)$ is the number of solutions

$$(V_0 \sqcup \mathcal{V}_{c'} \sqcup \{v_{c'+1}, \dots, v_{i''+1}\}) \sqcup (Y_0 \sqcup \mathcal{Y}_{c'} \sqcup \{y_{c'+1}, \dots, y_{i''+1}\})$$

that satisfies:

- (1) $(V_0 \sqcup \mathcal{V}_{c'} \sqcup \{v_{c'+1}, \ldots, v_{i''}\}) \sqcup (Y_0 \sqcup \mathcal{Y}_{c'} \sqcup \{y_{c'+1}, \ldots, y_{i''}\})$ satisfies $h_C(i)$;
- (2) $v_{i''+1} \oplus y_{i''+1} = \lambda_{i''+1};$
- $\begin{array}{c} (3) \quad v_{i''+1} \notin V_0 \sqcup \mathcal{V}_{c'} \sqcup \{v_{c'+1}, \dots, v_{i''}\}; \\ (4) \quad y_{i''+1} \notin Y_0 \sqcup \mathcal{Y}_{c'} \sqcup \{y_{c'+1}, \dots, y_{i''}\}. \end{array}$

Clearly, there are in total $Nh_C(i)$ solutions that only comply with the conditions (1) and (2), which do not take the conditions (3) and (4) into account. On the other hand, condition (3) excludes $|V_0 \cup \mathcal{V}_{c'}| + i$ possible solutions for $v_{i''+1}$. While condition (4) excludes $|Y_0 \cup \mathcal{Y}_{c'}| + i$ possible solutions for $y_{i''+1}$. Further, there are exactly $\delta_{i''+1}$ previous equations that share the same λ value with the equation in the i''-th component. By the principle of inclusion-exclusion, we obtain

$$h_C(i+1) \ge \left(N - (p + |\mathcal{V}_{c'}|) - i - (p + |\mathcal{Y}_{c'}|) - i + \delta_{i''+1}\right) h_C(i).$$

Then for $0 \le i \le c_5 - 1$ and $i'' = c_1 + c_2 + c_3 + c_4 + i$, we have

$$\begin{split} &\frac{h_C(i+1)(N-\delta_{i''+1})}{h_C(i)(N-(p+|\mathcal{V}_{c'}|)-i)(N-(p+|\mathcal{Y}_{c'}|)-i)} \\ \geq &\frac{(N-\delta_{i''+1})\Big(N-(p+|\mathcal{V}_{c'}|)-(p+|\mathcal{Y}_{c'}|)-2i+\delta_{i''+1}\Big)}{(N-(p+|\mathcal{V}_{c'}|)-i)(N-(p+|\mathcal{Y}_{c'}|)-i)} \\ \geq &\frac{(N-\delta_{i''+1})\Big(N-(p+|\mathcal{V}_{c'}|)-(p+|\mathcal{Y}_{c'}|)-2i+\delta_{i''+1}\Big)}{N^2-(2p+|\mathcal{V}_{c'}|+2i)N+(p+|\mathcal{V}_{c'}|+i)(p+|\mathcal{Y}_{c'}|+i)} \\ \geq &1-\frac{(p+|\mathcal{V}_{c'}|+i)(p+|\mathcal{Y}_{c'}|+i)+\delta_{i''+1}^2}{N^2-(2p+|\mathcal{V}_{c'}|+|\mathcal{Y}_{c'}|+2i)N+(p+|\mathcal{V}_{c'}|+i)(p+|\mathcal{Y}_{c'}|+i)} \\ \geq &1-\frac{(p+q_m)^2+\delta_{i''+1}^2}{N^2/2} \\ \geq &1-\frac{2(p+q_m)^2+2\delta^2}{N^2}\,, \end{split}$$

using that $|\mathcal{V}_{c'}| + i, |\mathcal{V}_{c'}| + i \leq q_m, 2p + |\mathcal{V}_{c'}| + |\mathcal{Y}_{c'}| + 2i \leq 2(p+q_m) \leq N/2$, and $\delta_{i''+1}^j \leq \delta$. Here we have $\delta = i \leq q_m$ in the case of prp-like settings, and $\delta = 0$ in the case of prf-like settings.

Since $q_5 = c_5$, we get

$$\frac{h_C(c_5)\prod_{i=c_1+c_2+c_3+c_4+1}^c(N-\delta_i)}{(N-p-q_2-c_3-q_4)_{q_5}(N-p-q_1-q_3-c_4)_{q_5}} \ge \prod_{i=c_1+c_2+c_3+c_4}^{c-1} \left(1-\frac{2(p+q_m)^2+2\delta^2}{N^2}\right) \ge 1-\sum_{i=c_1+c_2+c_3+c_4}^{c-1} \left(\frac{2(p+q_m)^2+2\delta^2}{N^2}\right) \ge 1-\frac{2q_m(p+q_m)^2+2q_m\delta^2}{N^2}.$$
(10)

Conclusion. Putting (11), (13), (14), and (15) together, we have

$$\begin{split} & \frac{h(\mathcal{G})\prod_{i=1}^{c}\prod_{j=1}^{\eta_{i}}(N-\delta_{i}^{j})}{(N-p)_{q_{2}+c_{3}+q_{4}+q_{5}}(N-p)_{q_{1}+q_{3}+c_{4}+q_{5}}} \\ &= \frac{h_{A}(c_{1}+c_{2})\prod_{i=1}^{c_{1}+c_{2}}\prod_{j=1}^{\eta_{i}}(N-\delta_{i}^{j})}{(N-p)_{q_{2}}(N-p)_{q_{1}}} \cdot \frac{h_{B}(c_{3}+c_{4})\prod_{i=c_{1}+c_{2}+1}^{c_{3}+c_{4}}\prod_{j=1}^{\eta_{i}}(N-\delta_{i}^{j})}{(N-p-q_{2})_{c_{3}+q_{4}}(N-p-q_{1})_{q_{3}+c_{4}}} \\ &\times \frac{h_{C}(c_{5})\prod_{i=c_{1}+c_{2}+c_{3}+c_{4}+1}(N-\delta_{i})}{(N-p-q_{2}-c_{3}-q_{4})_{q_{5}}(N-p-q_{1}-q_{3}-c_{4})_{q_{5}}} \\ &= 1 - \frac{2(p+q_{m})^{2}}{N^{2}} \Big(\sum_{i=c_{1}+c_{2}}^{c_{1}+c_{2}+c_{3}+c_{4}-1}(\eta_{i+1}^{2}+\eta_{i+1}) + q_{m}\Big) - \frac{\sum_{i=1}^{c_{1}+c_{2}}\delta}{N} - \frac{3q_{m}\delta^{2}}{N^{2}} \,. \end{split}$$

В Proof of Theorem 2

Since the isolated vertices in \mathcal{I} are already well defined, we focus only on the vertices in $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and \mathcal{D} . For readability, we use $N = 2^n, \mathcal{G}$ instead of $\mathcal{G}(\mathcal{E}^p)$, $\mathcal{G}^{=}$ instead of $\mathcal{G}(\mathcal{E}_m^p)$, and \mathcal{G}^{\neq} instead of $\mathcal{G}(\mathcal{E}_a)$. Let $i = 1, \ldots, c + c_6 + c_7$ with $c = c_1 + c_2 + c_3 + c_4 + c_5$. Hence, we have that \mathcal{A}_1 is the first component and \mathcal{D}_{c_7} is the last component, which is an isolated y vertex in $\mathcal{G}^=$.

- Let ξ_i be the size (number of vertices) of the *i*-th component, and $\eta_i = \xi_i 1$ (number of edges in the *i*-th component).
- for $j \in [\eta_i]$ and $r = \sum_{k=1}^{i-1} \eta_k + j$,
 - λ^j_i = λ^r (λ value corresponding to the *j*-th equation of *i*-th component).
 δ^j_i := μ(λ^{r-1}, λ^j_i) with δ¹₁ = 0 for prp-like settings.

 - $\delta_i^j := 0$ for prf-like settings.

When the *i*-th component only contains one edge, we drop the superscript j, and use λ_i and δ_i instead of λ_i^j and δ_i^j .

- For each component, where one part consists of the vertices in \mathcal{V}^{p} and the other vertices in \mathcal{Y}^{p} ; we denote the two parts respectively by V_i and Y_i , and we have $\xi_i = |V_i| + |Y_i|$. We denote by v_i^j (resp., y_i^j) the *j*-th vertex in V_i (resp., Y_i). When V_i (resp., Y_i) only contains one vertex, we will drop the superscript, and write v_i (resp., y_i).
- Let a_i^v be the number of λ' -labeled edges that connect a v vertex in the *i*-th component with a y vertex in one of the previous components. Similarly, let a_i^y be the number of λ' -labeled edges that connect a y vertex in the ith component with a v vertex in one of the previous components. We have $a_i = a_i^v + a_i^y$ and $q_a = \sum_{i=1}^{c+c_6+c_7} a_i$.

Recall there are exactly p colliding v vertices and p colliding y vertices in \mathcal{G} ; and that q_1, q_2, q_3, q_4 , and q_5 are the number of equations (edges) in $\mathcal{A}_1 \sqcup \cdots \sqcup \mathcal{A}_{c_1}$, $\mathcal{A}_{c_1+1} \sqcup \cdots \sqcup \mathcal{A}_{c_1+c_2}, \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_{c_3}, \mathcal{B}_{c_3+1} \sqcup \cdots \sqcup \mathcal{B}_{c_3+c_4}, \text{ and } \mathcal{C}, \text{ respectively.}$

Colliding Components. Recall that for a good graph, there is only one colliding vertex in each colliding component (Definition 5). For $i = 1, \ldots, c_1, A_i$ contains one colliding v vertex and η_i unknown y vertices ($\eta_i = 1$ in the case of isolated components). For $i = c_1 + 1, \ldots, c_1 + c_2, A_i$ contains one colliding y vertex and η_i unknown v vertices ($\eta_i = 1$ in the case of isolated components).

For $i = 1, \ldots, c_1 + c_2$, let $h_A(i)$ be the number of solutions to $\mathcal{I} \sqcup \mathcal{A}_1 \sqcup \cdots \sqcup \mathcal{A}_i$. Then we have that $h_A(i) = 1$ for $1 \leq i \leq c_1 + c_2$. For $1 \leq i \leq c_1$, let $v_i^* \in V_i$ be the colliding vertex of \mathcal{A}_i such that the value of this vertex is well defined, then the other η_i unknown y vertices in \mathcal{A}_i are uniquely determined, since each y vertex in \mathcal{A}_i is connected with a λ -labeled edge to v_i^* . Similarly, for $c_1 + 1 \leq i \leq c_1 + c_2$, let $y_i^* \in Y_i$ be the colliding vertex of \mathcal{A}_i such that the value of this vertex is well defined, then the other η_i unknown v vertices in \mathcal{A}_i are uniquely determined, since each y is some that the other η_i unknown v vertices in \mathcal{A}_i are uniquely determined, since each v vertex in \mathcal{A}_i is connected with a λ -labeled edge to y_i^* . Then, we have

$$\frac{h_A(c_1+c_2)\prod_{i=1}^{c_1+c_2}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{(N-p)_{q_2}(N-p)_{q_1}} \ge \frac{\prod_{i=1}^{c_1+c_2}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{N^{q_1+q_2}} \\
\ge \frac{N^{\sum_{i=1}^{c_1+c_2}\eta_i} - \sum_{i=1}^{c_1+c_2}\sum_{j=1}^{\eta_i}\delta_i^j N^{\sum_{i=1}^{c_1+c_2}\eta_i-1}}{N^{q_1+q_2}} \\
\ge \frac{N^{\sum_{i=1}^{c_1+c_2}\eta_i}(1-\sum_{i=1}^{c_1+c_2}\sum_{j=1}^{\eta_i}\delta_i^j/N)}{N^{q_1+q_2}} \\
\ge 1 - \frac{\sum_{i=1}^{c_1+c_2}\delta}{N}, \quad (11)$$

using $\sum_{i=1}^{c_1+c_2} \eta_i = q_1 + q_2$, and $\sum_{j=1}^{\eta_i} \delta_i^j \leq \delta$. Here we have $\delta = q_m$ in the case of prp-like settings, and $\delta = 0$ in the case of prf-like settings.

For the further analysis, we define

$$\mathcal{V}_i = \bigsqcup_{c_1 + 1 \le i \le c_1 + c_2} V_i \sqcup \bigsqcup_{c_1 + c_2 + 1 \le k \le i} V_k$$
$$\mathcal{Y}_i = \bigsqcup_{1 \le i \le c_1} Y_i \sqcup \bigsqcup_{c_1 + c_2 + 1 \le k \le i} Y_k.$$

Star Components. It can be easily seen that the analysis of the v- \star components and that of the y- \star components are symmetric, so we only perform the analysis for the v- \star components. Let $i = 1, \ldots, c_3$ and $i' = c_1 + c_2 + i$, and let $h_B(i)$ be the number of solutions to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_i$. Note that $h_B(0) = h(c_1 + c_2)$. Our goal is to derive a recursive formula for $h_B(i+1)$ that depends on $h_B(i)$, such that a lower bound can be found for the expression $h_B(i+1)/h_B(i)$. In order to do that, we fix a solution to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_i$. If we fix the single vertex $v_{i'+1} \in V_{i'+1}$ and assign any value to $v_{i'+1}$, then the other $\eta_{i'+1}$ unknowns $y_{i'+1}^{j}$ in the (i' + 1)-th component (\mathcal{B}_{i+1}) are uniquely determined, since each $y_{i'+1}^{j}$ vertex in Y_{i+1} is connected with a λ -labeled edge to $v_{i'+1}$, for $j \in [\eta_{i'+1}]$.

Then, $h_B(i+1)$ is the number of solutions

$$(V_0 \sqcup \mathcal{V}_{i'} \sqcup v_{i'+1}) \sqcup \left(Y_0 \sqcup \mathcal{Y}_{i'} \sqcup \{y_{i'+1}^1, \dots, y_{i'+1}^{\eta_{i'+1}}\}\right)$$

that satisfies:

- (1) $(V_0 \sqcup \mathcal{V}_{i'}) \sqcup (Y_0 \sqcup \mathcal{Y}_{i'})$ satisfies $h_B(i)$;
- (2) $v_{i'+1} \oplus y_{i'+1}^j = \lambda_{i'+1}^j$ for $j \in [\eta_{i'+1}];$ (3) $v_{i'+1} \notin (V_0 \sqcup \mathcal{V}_{i'});$
- (4) $y_{i'+1}^{j} \notin (Y_0 \sqcup \mathcal{Y}_{i'}), \text{ for } j \in [\eta_{i'+1}];$
- (5) $v_{i'+1}$ satisfies $a_{i'+1}^v$ non-equations to the y vertices in the previous components;
- (6) $y_{i'+1}^j$ satisfies $a_{i'+1}^j$ non-equations to the v vertices in the previous components, for $j \in [\eta_{i'+1}]$.

Clearly, there are in total $Nh_B(i)$ solutions that only comply with the conditions (1) and (2), which do not take the conditions (3), (4), (5), and (6) into account. On the other hand, condition (3) excludes $|V_0 \sqcup \mathcal{V}_{i'}|$ possible solutions for $v_{i'+1}$. While condition (4) excludes $|Y_0 \sqcup \mathcal{Y}_{i'}|$ possible solutions for each $y_{i'+1}^j$, hence, (4) excludes at most $|Y_0 \sqcup \mathcal{Y}_{i'}| \eta_{i'+1}$ possible solutions. Further, there are exactly $\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^{j}$ previous equations that share the same λ value with some equation in \mathcal{B}_{i+1} . Moreover, condition (5) excludes $a_{i'+1}^v$ possible solutions for $v_{i'+1}$, and condition (6) excludes $a_{i'+1}^y$ possible solutions for $y_{i'+1}^j$'s. By the principle of inclusion-exclusion, we obtain

$$h_B(i+1) \ge \left(N - (p + |\mathcal{V}_{i'}|) - a_{i'+1}^v - (p + |\mathcal{Y}_{i'}|)\eta_{i'+1} - a_{i'+1}^y + \sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j\right)h_B(i)$$

Then for $0 \le i \le c_3 - 1$ and $i' = c_1 + c_2 + i$, we have

$$\frac{h_B(i+1)\prod_{j=1}^{\eta_{i'+1}}(N-\delta_{i'+1}^j)}{h_B(i)(N-(p+|\mathcal{V}_{i'}|))(N-(p+|\mathcal{Y}_{i'}|))_{\eta_{i'+1}}} \\
\geq \frac{\prod_{j=1}^{\eta_{i'+1}}(N-\delta_{i'+1}^j)\Big(N-(p+|\mathcal{V}_{i'}|)-(p+|\mathcal{Y}_{i'}|)\eta_{i'+1}+\sum_{k=1}^{\eta_{i'+1}}\delta_{i'+1}^j-a_{i'+1}\Big)}{(N-(p+|\mathcal{V}_{i'}|))(N-(p+|\mathcal{Y}_{i'}|))_{\eta_{i'+1}}} \\
\geq 1-\frac{B-C}{(N-(p+|\mathcal{V}_{i'}|))(N-(p+|\mathcal{Y}_{i'}|))_{\eta_{i'+1}}},$$
(12)

where

$$B = (N - (p + |\mathcal{V}_{i'}|))(N - (p + |\mathcal{Y}_{i'}|))_{\eta_{i'+1}}$$

$$C = \prod_{j=1}^{\eta_{i'+1}} (N - \delta_{i'+1}^j) \left(N - (p + |\mathcal{V}_{i'}|) - (p + |\mathcal{Y}_{i'}|)\eta_{i'+1} + \sum_{k=1}^{\eta_{i'+1}} \delta_{i'+1}^j - a_{i'+1} \right).$$

We need both a lower and an upper bound for B. Using the facts that $|\mathcal{V}_{i'}|, |\mathcal{Y}_{i'}| + \eta_{i'+1} < q_m$, and $\xi_{\max}(p+q_m) < 2^{n-1}$, we get $B \ge N^{(\eta_{i'+1}+1)}/2$. We now derive an upper bound for B.

$$\begin{split} B &\leq (N - (p + |\mathcal{V}_{i'}|))(N - (p + |\mathcal{Y}_{i'}|))^{\eta_{i'+1}} \\ &\leq (N - (p + |\mathcal{V}_{i'}|))\left(N^{\eta_{i'+1}} - \eta_{i'+1}(p + |\mathcal{Y}_{i'}|)N^{\eta_{i'+1}-1} + \eta^2_{i'+1}(p + |\mathcal{Y}_{i'}|)^2 N^{\eta_{i'+1}-2}\right) \\ &\leq N^{\eta_{i'+1}+1} - \eta_{i'+1}(p + |\mathcal{Y}_{i'}|)N^{\eta_{i'+1}} + \eta^2_{i'+1}(p + |\mathcal{Y}_{i'}|)^2 N^{\eta_{i'+1}-1} \\ &- (p + |\mathcal{V}_{i'}|)N^{\eta_{i'+1}} + \eta_{i'+1}(p + |\mathcal{V}_{i'}|)(p + |\mathcal{Y}_{i'}|)N^{\eta_{i'+1}-1}. \end{split}$$

We now derive a lower bound for C.

$$\begin{split} C &\geq \Big(N^{\eta_{i'+1}} - \sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j N^{\eta_{i'+1}-1} \Big) \Big(N - (p + |\mathcal{V}_{i'}|) - (p + |\mathcal{Y}_{i'}|) \eta_{i'+1} + \sum_{k=1}^{\eta_{i'+1}} \delta_{i'+1}^j - a_{i'+1} \Big) \\ &\geq N^{\eta_{i'+1}+1} - (p + |\mathcal{V}_{i'}|) N^{\eta_{i'+1}} - \eta_{i'+1} (p + |\mathcal{Y}_{i'}|) N^{\eta_{i'+1}} \\ &- \Big(\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j \Big)^2 N^{\eta_{i'+1}-1} - a_{i'+1} N^{\eta_{i'+1}} \,. \end{split}$$

On substituting the bounds of B and C in (12), we get

$$\begin{aligned} (12) &\geq 1 - \frac{\eta_{i'+1}^2 (p + |\mathcal{Y}_{i'}|)^2 N^{\eta_{i'+1}-1} + \eta_{i'+1} (p + |\mathcal{V}_{i'}|) (p + |\mathcal{Y}_{i'}|) N^{\eta_{i'+1}-1}}{N^{\eta_{i'+1}+1}/2} \\ &- \frac{\left(\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j\right)^2 N^{\eta_{i'+1}-1} + a_{i'+1} N^{\eta_{i'+1}}}{N^{\eta_{i'+1}+1}/2} \\ &\geq 1 - \frac{\eta_{i'+1}^2 (p + q_m)^2 N^{\eta_{i'+1}-1} + \eta_{i'+1} (p + q_m)^2 N^{\eta_{i'+1}-1}}{N^{\eta_{i'+1}+1}/2} \\ &- \frac{\left(\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j\right)^2 N^{\eta_{i'+1}-1} + a_{i'+1} N^{\eta_{i'+1}}}{N^{\eta_{i'+1}+1}/2} \\ &\geq 1 - \frac{2(p + q_m)^2 (\eta_{i'+1}^2 + \eta_{i'+1}) + 2\delta^2}{N^2} - \frac{2a_{i'+1}}{N}, \end{aligned}$$

using the fact that $|\mathcal{V}_{i'}| = q_2 + c_3 + q_4 \leq q_m, |\mathcal{Y}_{i'}| = q_1 + q_3 + c_4 \leq q_m$, and $\sum_{j=1}^{\eta_{i'+1}} \delta_{i'+1}^j \leq \delta$. Here we have $\delta = q_m$ in the case of prp-like settings $(\lambda_{i'+1}^j \text{ can occur at most once in any component})$, and $\delta = 0$ in the case of prf-like settings.

We consider the $c_3 v \star$ and the $c_4 y \star$ components together

$$\frac{h_B(c_3+c_4)\prod_{i=c_1+c_2+1}^{c_1+c_2+c_3+c_4}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{(N-p-q_2)_{c_3+q_4}(N-p-q_1)_{q_3+c_4}} \\
\geq \prod_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} \left(1-\frac{2(p+q_m)^2(\eta_{i+1}^2+\eta_{i+1})+2\delta^2}{N^2}-\frac{2a_{i+1}}{N}\right) \\
\geq 1-\sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} \left(\frac{2(p+q_m)^2(\eta_{i+1}^2+\eta_{i+1})+2\delta^2}{N^2}-\frac{2a_{i+1}}{N}\right) \\
\geq 1-\frac{2(p+q_m)^2\sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1}(\eta_{i+1}^2+\eta_{i+1})+q_m\delta^2}{N^2}-\frac{2\sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1}a_{i+1}}{N}\right) \\$$
(13)

using the fact that $c_3 + c_4 \le (q_3 + q_4)/2 \le q_m/2$.

Isolated components. We now consider the components of the graph that contain exactly two vertices. Let $c' = c_1 + c_2 + c_3 + c_4$. For $i = 1, \ldots, c_5$ and $i'' = c_1 + c_2 + c_3 + c_4 + i$, we will write

$$\mathcal{C}_i: v_{i^{\prime\prime}} \stackrel{\lambda_{i^{\prime\prime}}}{-} y_{i^{\prime\prime}}$$

where $v_{i''} \in V_{i''}$ and $y_{i''} \in Y_{i''}$. Let $h_C(i)$ be the number of solutions to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B} \sqcup \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$. Note that $h_C(0) = h(c_1 + c_2 + c_3 + c_4)$. Our goal is to derive a recursive formula for $h_C(i+1)$ that depends on $h_C(i)$, such that a lower bound can be found for the expression $h_C(i+1)/h_C(i)$. In order to do that, we fix a solution to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B} \sqcup \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$. Then, $h_C(i+1)$ is the number of solutions

$$(V_0 \sqcup \mathcal{V}_{c'} \sqcup \{v_{c'+1}, \dots, v_{i''+1}\}) \sqcup (Y_0 \sqcup \mathcal{Y}_{c'} \sqcup \{y_{c'+1}, \dots, y_{i''+1}\})$$

that satisfies:

- (1) $(V_0 \sqcup \mathcal{V}_{c'} \sqcup \{v_{c'+1}, \dots, v_{i''}\}) \sqcup (Y_0 \sqcup \mathcal{Y}_{c'} \sqcup \{y_{c'+1}, \dots, y_{i''}\})$ satisfies $h_C(i)$;
- (2) $v_{i''+1} \oplus y_{i''+1} = \lambda_{i''+1};$
- (3) $v_{i''+1} \notin V_0 \sqcup \mathcal{V}_{c'} \sqcup \{v_{c'+1}, \dots, v_{i''}\};$
- (4) $y_{i''+1} \notin Y_0 \sqcup \mathcal{Y}_{c'} \sqcup \{y_{c'+1}, \dots, y_{i''}\};$
- (5) $v_{i''+1}$ satisfies $a_{i''+1}^v$ non-equations to the y vertices in the previous components;
- (6) $y_{i''+1}$ satisfies $a_{i''+1}^y$ non-equations to the v vertices in the previous components.

Clearly, there are in total $Nh_C(i)$ solutions that only comply with the conditions (1) and (2), which do not take the conditions (3), (4), (5), and (6) into account. On the other hand, condition (3) excludes $|V_0 \cup \mathcal{V}_{c'}| + i$ possible solutions for $v_{i''+1}$. While condition (4) excludes $|Y_0 \cup \mathcal{V}_{c'}| + i$ possible solutions for $y_{i''+1}$. Further, there are exactly $\delta_{i''+1}$ previous equations that share the same λ value with the equation in the i''-th component. Moreover, condition (5) excludes $a_{i''+1}^v$ possible solutions for $v_{i''+1}$, and condition (6) excludes $a_{i''+1}^y$ possible solutions for $y_{i''+1}$. By the principle of inclusion-exclusion, we obtain

$$h_C(i+1) \ge \left(N - (p + |\mathcal{V}_{c'}|) - i - a_{i''+1}^v - (p + |\mathcal{Y}_{c'}|) - i - a_{i''+1}^y + \delta_{i''+1}\right) h_C(i).$$

Then for $0 \le i \le c_5 - 1$ and $i'' = c_1 + c_2 + c_3 + c_4 + i$, we have

$$\begin{split} &\frac{h_C(i+1)(N-\delta_{i''+1})}{h_C(i)(N-(p+|\mathcal{V}_{c'}|)-i)(N-(p+|\mathcal{Y}_{c'}|)-i)} \\ \geq \frac{(N-\delta_{i''+1})\Big(N-(p+|\mathcal{V}_{c'}|)-(p+|\mathcal{Y}_{c'}|)-2i+\delta_{i''+1}-(a_{i''+1}^v+a_{i''+1}^y)\Big)}{(N-(p+|\mathcal{V}_{c'}|)-i)(N-(p+|\mathcal{Y}_{c'}|)-i)} \\ \geq \frac{(N-\delta_{i''+1})\Big(N-(p+|\mathcal{V}_{c'}|)-(p+|\mathcal{Y}_{c'}|)-2i+\delta_{i''+1}-a_{i''+1}\Big)}{N^2-(2p+|\mathcal{V}_{c'}|+i)(p+|\mathcal{Y}_{c'}|+i)+\delta_{i''+1}^2+a_{i''+1}N} \\ \geq 1-\frac{(p+|\mathcal{V}_{c'}|+i)(p+|\mathcal{Y}_{c'}|+i)+\delta_{i''+1}^2+a_{i''+1}N}{N^2-(2p+|\mathcal{V}_{c'}|+|\mathcal{Y}_{c'}|+2i)N+(p+|\mathcal{V}_{c'}|+i)(p+|\mathcal{Y}_{c'}|+i)} \\ \geq 1-\frac{(p+q_m)^2+\delta_{i''+1}^2+a_{i''+1}N}{N^2/2} \\ \geq 1-\frac{2(p+q_m)^2+2\delta^2}{N^2}-\frac{2a_{i''+1}}{N}, \end{split}$$

using that $|\mathcal{V}_{c'}| + i$, $|\mathcal{V}_{c'}| + i \leq q_m$, $2p + |\mathcal{V}_{c'}| + |\mathcal{Y}_{c'}| + 2i \leq 2(p + q_m) \leq N/2$, and $\delta_{i''+1}^j \leq \delta$. Here we have $\delta = i \leq q_m$ in the case of prp-like settings, and $\delta = 0$ in the case of prf-like settings.

Since $q_5 = c_5$, we get

$$\frac{h_C(c_5)\prod_{i=c_1+c_2+c_3+c_4+1}^c (N-\delta_i)}{(N-p-q_2-c_3-q_4)_{q_5}(N-p-q_1-q_3-c_4)_{q_5}} \\
\geq \prod_{i=c_1+c_2+c_3+c_4}^{c-1} \left(1-\frac{2(p+q_m)^2+2\delta^2}{N^2}-\frac{2a_{i+1}}{N}\right) \\
\geq 1-\sum_{i=c_1+c_2+c_3+c_4}^{c-1} \left(\frac{2(p+q_m)^2+2\delta^2}{N^2}-\frac{2a_{i+1}}{N}\right) \\
\geq 1-\frac{2q_m(p+q_m)^2+2q_m\delta^2}{N^2}-\frac{2\sum_{i=c_1+c_2+c_3+c_4}^{c-1}a_{i+1}}{N}.$$
(14)

Isolated Vertices. Finally, we still need to consider the isolated vertices that are only connected with λ' -labeled edges. So these are isolated vertices in $\mathcal{G}^=$ where the values of the vertices are unknown.

We only consider the case for the isolated v vertices, the analysis of isolated y vertices can be performed in a similar way. For $i = 1, \ldots, c_6$ and $i^* = c + i$, let $h_D(i)$ be the number of solutions to $\mathcal{I} \sqcup \mathcal{A} \sqcup \mathcal{B} \sqcup \mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i$, where \mathcal{D}_i

is an isolated vertex in $\mathcal{G}^{=}$, which is connected to λ' -labeled edges in \mathcal{G}^{\neq} . Note that $h_D(0) = h(c)$ and $h_D(c_6 + c_7) = h(\mathcal{G})$. Then, $h_D(i+1)$ is the number of solutions $V_0 \sqcup \mathcal{V}_c \sqcup \{v'_{c+1}, \ldots, v'_{i^*+1}\}$, that satisfies:

- (1) $V_0 \sqcup \mathcal{V}_c \sqcup \{v'_{c+1}, \dots, v'_{i^*}\}$ satisfies $h_D(i)$; (2) $v'_{i^*+1} \notin V_0 \sqcup \mathcal{V}_c \sqcup \{v'_{c+1}, \dots, v'_{i^*}\}$; (3) v'_{i^*+1} should satisfy $a^{v}_{i^*+1}$ non-equations to the y vertices.

Clearly, there are in total $Nh_D(i)$ solutions that only comply with the condition (1), which do not take the conditions (2) and (3) into account. On the other hand, condition (2) excludes $|V_0 + \mathcal{V}_c| + i$ possible solutions, and condition (5) excludes $a_{i^*+1}^v$ possible solutions. Therefore, we have $h_D(c_6) \ge \prod_{i=0}^{c_6-1} (N-p-1)^{i_6-1}$ $|\mathcal{V}_c| - i - a_{i^*+1}^v)$, which means

$$\frac{h_D(c_6)}{(N-p-q_2-c_3-q_4-q_5)_{c_6}} \ge \prod_{i=0}^{c_6-1} \frac{N-p-|\mathcal{V}_c|-i-a_{i^*+1}^v}{N-p-|\mathcal{V}_c|-i}$$
$$\ge \prod_{i=0}^{c_6-1} \left(1 - \frac{a_{i^*+1}^v}{N-p-|\mathcal{V}_c|-i}\right)$$
$$\ge 1 - \frac{2\sum_{i^*=c}^{c_4-c_6-1} a_{i^*+1}^v}{N}.$$

We consider the c_6 isolated v vertices and the c_7 isolated y vertices in \mathcal{D} together.

$$\frac{h_D(c_6+c_7)}{(N-p-q_2-c_3-q_4-q_5)_{c_6}(N-p-q_1-q_3-c_4-q_5)_{c_7}} \ge \left(1-\frac{2\sum_{i=c}^{c+c_6-1}a_{i+1}^v}{N}\right)\left(1-\frac{2\sum_{i=c+c_6}^{c+c_6+c_7-1}a_{i+1}^y}{N}\right).$$
 (15)

Conclusion. Putting (11), (13), (14), and (15) together, and using $\sum_{i=c_1+c_2}^{c+c_6+c_7} a_i \leq 1$ $\sum_{i=1}^{c+c_6+c_7} a_i \leq q_a$. We have

$$\begin{split} & \frac{h(\mathcal{G})\prod_{i=1}^{c+c_6+c_7}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{(N-p)_{q_2+c_3+q_4+q_5+c_6}(N-p)_{q_1+q_3+c_4+q_5+c_7}} \\ &= \frac{h_A(c_1+c_2)\prod_{i=1}^{c_1+c_2}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{(N-p)_{q_2}(N-p)_{q_1}}\cdot\frac{h_B(c_3+c_4)\prod_{i=c_1+c_2+1}^{c_3+c_4}\prod_{j=1}^{\eta_i}(N-\delta_i^j)}{(N-p-q_2)_{c_3+q_4}(N-p-q_1)_{q_3+c_4}} \\ &\times \frac{h_C(c_5)\prod_{i=c_1+c_2+c_3+c_4+1}^{c_1}(N-\delta_i)}{(N-p-q_2-c_3-q_4)_{q_5}(N-p-q_1-q_3-c_4)_{q_5}} \\ &\times \frac{h_D(c_6+c_7)}{(N-p-q_2-c_3-q_4-q_5)_{c_6}(N-p-q_1-q_3-c_4-q_5)_{c_7}} \\ &= 1-\frac{2(p+q_m)^2}{N^2} \Big(\sum_{i=c_1+c_2}^{c_1+c_2+c_3+c_4-1} (\eta_{i+1}^2+\eta_{i+1})+q_m\Big) - \frac{\sum_{i=1}^{c_1+c_2}\delta}{N} - \frac{3q_m\delta^2}{N^2} - \frac{2q_a}{N} \,. \end{split}$$

C Proof of Lemma 4

The proof generalizes over the single-user proof of Cogliati et al. [15]. We want to bound the probability $\Pr[X_{\mathcal{P}} \in \mathcal{T}_{bad}]$ that an ideal world transcript τ satisfies one of the bad events. Therefore, by the union bound, the probability that $X_{\mathcal{P}} \in \mathcal{T}_{bad}$ can be bounded as

$$\Pr[X_{\mathcal{P}} \in \mathcal{T}_{\mathrm{bad}}] \le \sum_{i} \Pr[\mathrm{bad}_{i}].$$

Events bad₁, bad₂, and bad₃. We consider the event bad₁. For any $(j, T, M, C) \in \tau_m$, $(u, v) \in \tau_1$, and $(x, y) \in \tau_2$, one has, by the fact that h_1^j and h_2^j are independently drawn,

$$\Pr[(h_1^j(T) = M \oplus u) \land (h_2^j(T) = C \oplus y)] = \epsilon^2$$

Hence, summing over the q possible (j, T, M, C), p possible (u, v), and p possible (x, y) yields

$$\Pr[\mathrm{bad}_1] \le q p^2 \epsilon^2 \le \sqrt{q} p \epsilon.$$

Similarly, for bad_2 and bad_3 , one obtains

$$\begin{aligned} &\Pr[\text{bad}_2] \leq q p^2 \epsilon^2 \leq \sqrt{q} p \epsilon \,, \\ &\Pr[\text{bad}_3] \leq q p^2 \epsilon^2 \leq \sqrt{q} p \epsilon \,. \end{aligned}$$

Event bad₄. For any $(j, T, M, C), (j', T', M', C'), (j'', T'', M'', C'') \in \tau_m$ with (j', T', M', C') distinct from (j, T, M, C) and from (j'', T'', M'', C''), one has, by the fact that $h_1^j, h_1^{j'}, h_2^{j'}$, and $h_2^{j''}$ are drawn independently,

$$\Pr[(h_1^j(T) \oplus h_1^{j'}(T') = M \oplus M') \land (h_2^{j'}(T) \oplus h_2^{j''}(T'') = C \oplus C'')] \le \epsilon^2.$$

Hence, summing over at most q^3 possible (j,T,M,C), (j',T',M',C'), (j'',T'',M'',C''), one obtains

$$\Pr[\text{bad}_4] \le q^3 \epsilon^2$$
.

Events bad₅ and bad₆. For any two distinct queries $(j, T, M, C) \neq (j', T', M', C') \in \tau_m$, one has, by the fact that h_1^j , $h_1^{j'}$, h_2^j , and $h_2^{j'}$ are drawn independently,

$$\Pr[(h_1^j(t) \oplus h_1^{j'}(T') = M \oplus M') \land (h_2^j(T) \oplus h_2^{j'}(T') = h_1^j(T) \oplus h_1^{j'}(T'))] \le \epsilon^2.$$

Hence, summing over the q(q-1)/2 possible pairs of $(j, T, M, C) \neq (j', T', M', C')$, we get

$$\Pr[\mathrm{bad}_5] \le \frac{q^2 \epsilon^2}{2} \le \frac{q^3 \epsilon^2}{2}.$$

Similarly,

$$\Pr[\text{bad}_6] \le \frac{q^2 \epsilon^2}{2} \le \frac{q^3 \epsilon^2}{2}$$

Events bad₇ and bad₈. For any two distinct queries $(j, T, M, C) \neq (j', T', M', C') \in \tau_m$ and any $(u, v) \in \tau_1$, one has, by the fact that h_1^j , h_2^j , and $h_2^{j'}$ are drawn independently,

$$\Pr[(h_1^j(T) = M \oplus u) \land (h_2^j(T) \oplus h_2^{j'}(T') = C \oplus C')] \le \epsilon^2.$$

Then, summing over q(q-1)/2 possible pairs of $(j, T, M, C) \neq (j', T', M', C')$, and p possible (u, v),

$$\Pr[\mathrm{bad}_7] \le \frac{q^2 p \epsilon^2}{2} \,.$$

Similarly,

$$\Pr[\mathrm{bad}_8] \le \frac{q^2 p \epsilon^2}{2} \,.$$

Events bad₉, bad₁₀, bad_{c1} and bad_{c2}. We will deal with the events bad₉ and bad_{c1} together, using the fact that

$$\Pr[\operatorname{bad}_9 \lor \operatorname{bad}_{c_1}] = \Pr[\operatorname{bad}_{c_1}] + \Pr[\operatorname{bad}_9 \land \neg \operatorname{bad}_{c_1}].$$

To upper bound $\Pr[\text{bad}_{c_1}]$, we see c_1 as a random variable over the random choice of h_1^j . First, note that we have,

$$\boldsymbol{E}[c_1] = \sum_{(j,T,M,C)\in\tau_m} \sum_{(u,v)\in\tau_1} \Pr[M \oplus h_1^j(T) = u] = qp\epsilon \,,$$

so that by Markov's inequality,

$$\Pr[\operatorname{bad}_{c_1}] \leq \sqrt{q}p\epsilon$$
.

For this bad event $\operatorname{bad}_9 \wedge \neg \operatorname{bad}_{c_1}$, we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $h_1^j = h_1^{j'}$. Fix any $h_1' \in \mathcal{H}$ such that, when $h_1^j = h_1'$, $c_1 < \sqrt{q}$, and fix any queries $(j,T,M,C) \neq (j',T',M',C') \in \tau_m, (u,v), (u',v') \in \tau_1$ such that $M \oplus h_1^j(T) = u$ and $M' \oplus h_1^j(T') = u'$. Note that since $c_1 < \sqrt{q}$, there are at most q/2 such tuple of queries. Then

$$\Pr[(h_1^j = h_1') \land (h_2^j(T) \oplus h_2^j(T') = v \oplus h_1^j(T) \oplus v' \oplus h_1^j(T'))] \le \frac{\epsilon}{|\mathcal{H}|}$$

and, by summing over every h_1^j such that $c_1 < \sqrt{q}$ and every such tuple of queries, one has

$$\Pr[\operatorname{bad}_9(a) \land \neg \operatorname{bad}_{c_1}] \le \frac{q\epsilon}{2} \le \frac{q^3\epsilon^2}{2}$$

2. For case (b), the keys h_1^j and $h_1^{j'}$ are generated independently of each other. Using the fact that h_1^j , $h_1^{j'}$, h_2^j , and $h_2^{j'}$ are drawn independently, then, summing over at most q(q-1)/2 possible (j, T, M, C), (j', T', M', C'), and p^2 possible (u, v), (u', v'), one obtains

$$\Pr[\operatorname{bad}_9(b) \land \neg \operatorname{bad}_{c_1}] = \Pr[\operatorname{bad}_9(b)] \le q^2 p^2 \epsilon^3 / 2 \le \sqrt{q} p \epsilon / 2.$$

Finally,

$$\Pr[\operatorname{bad}_9 \lor \operatorname{bad}_{c_1}] \le \sqrt{q}p\epsilon + \frac{q^3\epsilon^2}{2} + \sqrt{q}p\epsilon/2.$$

Similarly,

$$\Pr[\operatorname{bad}_{10} \lor \operatorname{bad}_{c_2}] \le \sqrt{q}p\epsilon + \frac{q^3\epsilon^2}{2} + \sqrt{q}p\epsilon/2.$$

Events bad₁₁ and bad₁₂. Note that since the hash keys for different users are generated independently, it is enough to focus on the single user when $h_1^{j_{i_1}} = \cdots = h_1^{j_{i_{\xi+1}}}$ and $h_2^{j_{i_1}} = \cdots = h_2^{j_{i_{\xi+1}}}$, as this case will yield the dominant term. For the events bad₁₁, since, $(T^{(j_i)}, M^{(j_i)}) \neq (T^{(j_k)}, M^{(j_k)})$ for all $i \neq k$, we can apply Lemma 2 with $a = \xi = 2^n/2(p+q)$ to get

$$\Pr[\text{bad}_{11}] \le \frac{8q^2(p+q)^2\epsilon}{2^{2n}}.$$

Similarly

$$\Pr[\text{bad}_{12}] \le \frac{8q^2(p+q)^2\epsilon}{2^{2n}}.$$

The result follows by an union bound over all conditions.

D Proof of Lemma 5

The bad transcript analysis relies on the sum-capture theorem proposed by Chen et al. [8]. Typically, a sum-capture theorem states that for a random subset Z of $\{0, 1\}^n$ of size q, the quantity

$$\mu(Z, A, B) = |\{(z, a, b) \in Z \times A \times B \colon z = a \oplus b\}|$$

is not much larger than $q |A| |B| / 2^n$ for any possible choice of A and B, except with negligible probability. In our setting, Z will consist of query-response tuples from a permutation, i.e. Z consists of values $u_i \oplus v_i$ where $\{(u_1, v_1), \ldots, (u_q, v_q)\}$ is a permutation transcript. For this case, Chen et al. [8] proved the following result. **Lemma 7 (Chen et al.** [8]). Let Γ be an invertible linear map on the \mathbb{F}_2 -vector space $\{0,1\}^n$. Let $\pi \stackrel{\$}{\leftarrow}$ Permn, let \mathcal{D} be some probabilistic algorithm making exactly q distinct two-sided adaptive queries to π . Let $Z = \{(u_1, v_1), \ldots, (u_q, v_q)\}$ be the transcript of the interaction of \mathcal{D} with π , which consists of $q \ge 1$ pairs such that either $v_i = \pi(u_i)$ or $u_i = \pi(v_i)$ for all $i = 1, \ldots, q$. For any two subsets $A, B \subseteq \{0, 1\}^n$, let

$$\mu(Z,A,B) = |\{((u,v),a,b) \in Z \times A \times B \colon u \oplus a = \Gamma(v \oplus b)\}| \ .$$

Then, for $9n \leq q \leq 2^{n-1}$, we have

$$\Pr\left[\mu(Z, A, B) \ge \frac{q |A| |B|}{2^n} + \frac{2q^2 \sqrt{|A| |B|}}{2^n} + 3\sqrt{nq |A| |B|}\right] \le \frac{2}{2^n}.$$

The proof generalizes over the single-user proof of Dutta et al. [22]. We want to bound the probability $\Pr[X_{\mathcal{P}} \in \mathcal{T}_{\text{bad}}]$ that an ideal world transcript τ satisfies one of the bad events. Therefore, by the union bound, the probability that $X_{\mathcal{P}} \in \mathcal{T}_{\text{bad}}$ can be bounded as

$$\Pr[X_{\mathcal{P}} \in \mathcal{T}_{\text{bad}}] \le \sum_{i} \Pr[\text{bad}_{i}].$$

Event bad₁. We can rewrite bad₁ as

$$M \oplus u = C \oplus x = K_1^j.$$

We denote

$$\Omega_1 = \left| \left\{ \left((j, M, C), (u, v), (x, y) \right) \mid M \oplus u = C \oplus y \right\} \right| \,.$$

Then, using Lemma 7, there are Ω_1 possible combinations of $M \oplus C$, u and y that satisfy bad₁. Note that in the ideal world, Ω_1 only depends on φ_j , π_1 , and π_2 . Ω_1 does not depend on K_1^j , which is drawn uniformly at random at the end of the interaction. Hence, for any $C_1 > 0$, we have

$$\Pr[\operatorname{bad}_1] \le \Pr[\Omega_1 \ge C_1] + \frac{C_1}{2^n}.$$

We thus set $C_1 = \frac{qp^2}{2^n} + \frac{2q^2p}{2^n} + 3p\sqrt{nq}$ and obtain

$$\Pr[\text{bad}_1] \le \frac{2}{2^n} + \frac{qp^2}{2^{2n}} + \frac{2q^2p}{2^{2n}} + \frac{3p\sqrt{nq}}{2^n} + \frac$$

Events bad₂, and bad₃. We consider the event bad₂. For any $(j, M, C) \in \tau_m$, $(u, v) \in \tau_1$, and $(x, y) \in \tau_2$, one has, by the fact that K_1^j and K_2^j are independently drawn,

$$\Pr[(K_1^j = M \oplus u) \land (K_2^j = M \oplus v \oplus y \oplus K_1^j)] = \frac{1}{2^{2n}}.$$

Hence, summing over the q possible (j, M, C), p possible (u, v), and p possible (x, y) yields

$$\Pr[\text{bad}_2] \le \frac{qp^2}{2^{2n}}$$

Similarly, one obtains

$$\Pr[\operatorname{bad}_3] \le \frac{qp^2}{2^{2n}} \,.$$

Event bad₄. Note that this event only holds when j, j' and j'' are all distinct (three distinct users). For any (j, M, C), (j', M', C'), $(j'', M'', C'') \in \tau_m$, one has, by the fact that K_1^j , $K_1^{j'}$, $K_2^{j'}$, and $K_2^{j''}$ are drawn independently,

$$\Pr[(K_1^j \oplus K_1^{j'} = M \oplus M') \land (K_2^{j'} \oplus K_2^{j''} = C \oplus C'')] \le \frac{1}{2^{2n}}.$$

Hence, summing over at most q^3 possible $(j,M,C), (j^\prime,M^\prime,C^\prime), (j^{\prime\prime},M^{\prime\prime},C^{\prime\prime}),$ one obtains

$$\Pr[\text{bad}_4] \le \frac{q^3}{2^{2n}}$$

Events bad₅ and bad₆. Note that this event only holds when $j \neq j'$ (distinct users). For any two distinct queries $(j, M, C) \neq (j', M', C') \in \tau_m$, one has, by the fact that $K_1^j, K_1^{j'}, K_2^j$, and $K_2^{j'}$ are drawn independently,

$$\Pr[(K_1^j \oplus K_1^{j'} = M \oplus M') \land (K_2^j \oplus K_2^{j'} = M \oplus M' \oplus K_1^j \oplus K_1^{j'})] \le \frac{1}{2^{2n}}$$

Hence, summing over the q(q-1)/2 possible pairs of $(j, M, C) \neq (j', M', C')$, we get

$$\Pr[\text{bad}_5] \le \frac{q^2}{2^{2n-1}} \le \frac{q^3}{2^{2n-1}}$$

Similarly,

$$\Pr[\text{bad}_6] \le \frac{q^2}{2^{2n-1}} \le \frac{q^2}{2^{2n-1}}$$

Events bad_7 and bad_8 . For the event bad_7 , we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $K_j = K_{j'}$. For any two distinct queries $(j, M, C) \neq (j', M', C') \in \tau_m$, one has, by the fact that K_1^j , and the output C and C' are drawn independently,

$$\Pr[(K_1^j = M \oplus u) \land (C = C')] \le \frac{1}{2^{2n}}.$$

Hence, summing over the q(q-1)/2 possible pairs of $(j, M, C) \neq (j', M', C')$, and p possible pairs of (u, v), we get

$$\Pr[\operatorname{bad}_7(a)] \le \frac{q^2 p}{2^{2n-1}}$$

2. For case (b), the keys K_j and $K_{j'}$ are generated independently of each other. The second equation holds with probability $1/2^n$. The rest of the analysis is identical to that of case (a).

Hence

$$\Pr[\operatorname{bad}_7] \le \frac{q^2 p}{2^{2n-1}} \,.$$

For the event bad₈, note that this event only holds when $j \neq j'$ (distinct users). The analysis is identical to that of bad₇(b).

$$\Pr[\text{bad}_8] \le \frac{q^2 p}{2^{2n-1}} \,.$$

Events bad₉ and bad₁₀. For this bad event bad₉, we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $K_1^j = K_1^{j'}$ and $K_2^j = K_2^{j'}$, the event becomes $M \oplus u = K_1^j \wedge M' \oplus u' = K_1^j \wedge v \oplus M = v' \oplus M'$. In order to bound this bad event, we need the following help event:

 $\mathbb{Z}: \theta = |\{\{(u,v), (u',v')\}: (u,v), (u',v') \in \tau_p, u \oplus v = u' \oplus v'\}| \ge \sqrt{p}.$

Then we have the following:

$$\Pr[\mathrm{bad}_9(a) \vee \mathbb{Z}] \le \Pr[\mathbb{Z}] + \Pr[\mathrm{bad}_9(a) \wedge \neg \mathbb{Z}]$$

To bound the probability of the event \mathbb{Z} , we define an indicator random variable \mathbb{I}_{ik} which is set to 1 if and only if $(u_i, v_i), (u_k, v_k) \in \tau_p$ such that $u_i \oplus v_i = u_k \oplus v_k$. Therefore, we have

$$\theta = \sum_{i,k} \mathbb{I}_{ik} \,.$$

Now, for a fixed i, k, we have $\Pr[\mathbb{I}_{ik} = 1] = 1/2^n$. This is due to the fact that either both of $(u_i, v_i), (u_k, v_k)$ are backward queries in which u_i, u_k are random values or at least one of them is a forward query (w.l.og we assume (u_k, v_k) is a forward query) in which v_j is random. Hence, using the linearity of expectation, we have

$$\boldsymbol{E}[\theta] = \sum_{ik} \boldsymbol{E}[\mathbb{I}_{ik}] = \sum_{ik} \Pr[\mathbb{I}_{ik} = 1] \le \frac{p^2}{2^n} \,.$$

Therefore, using Markov's inequality, we have

$$\Pr[\mathbb{Z}] \le \frac{\boldsymbol{E}[\theta]}{\sqrt{p}} \le \frac{p^{3/2}}{2^n}$$

Now, we bound the probability of $\operatorname{bad}_9 \wedge \neg \mathbb{Z}$. To bound the event, for a fixed pair of $(j, M, C), (j', M', C') \in \tau_m$ and for a fixed pair of $(u, v), (u', v') \in \tau_1$, the probability that the event holds is $1/2^n$ due to the randomness of K_1^j . The number of (u, v), (u', v') satisfies the the event is at most \sqrt{p} . Moreover, the number of choices for (j, M, C) is q which makes the choice for (j', M', C') at most 1, as choosing an (j, M, C) determines (j', M', C'), namely, $M' = u' \oplus u \oplus x$. Hence,

$$\Pr[\mathrm{bad}_9(a) \land \neg \mathbb{Z}] \le \frac{q\sqrt{p}}{2^n}$$

2. For case (b), the keys K_1^j , $K_1^{j'}$, K_2^j , $K_2^{j'}$ are generated independently of each other. Using the fact that K_1^j , $K_1^{j'}$, K_2^j , $K_2^{j'}$ are drawn independently, then, summing over at most q(q-1)/2 possible (j, M, C), (j', M', C'), and p^2 possible (u, v), (u', v'), one obtains

$$\Pr[\text{bad}_9(b)] \le \frac{q^2 p^2}{2^{3n-1}} \le \frac{q p^2}{2^{2n-1}}.$$

Finally,

$$\Pr[\text{bad}_9 \vee \mathbb{Z}] \le \frac{q\sqrt{p}}{2^n} + \frac{p^{3/2}}{2^n} + \frac{qp^2}{2^{2n-1}}$$

Similarly,

$$\Pr[\operatorname{bad}_{10} \vee \mathbb{Z}] \le \frac{q\sqrt{p}}{2^n} + \frac{p^{3/2}}{2^n} + \frac{qp^2}{2^{2n-1}}.$$

Event bad₁₁. Note that since the keys for different users are generated independently of each other, it is sufficient to consider the single user case when $K_2^{j_{i_1}} = \cdots = K_2^{j_{i_{\xi+1}}}$. Using the fact that in the ideal world the randomness in the transcript τ_m is in the values $C_{i_1}, \ldots, C_{i_{\xi+1}} \stackrel{\$}{\leftarrow} \{0,1\}^n$. We obtain

$$\Pr[\mathrm{bad}_{11}] = \frac{\binom{q}{\xi+1}}{2^{n\xi}}.$$

The result follows by an union bound over all conditions.

E Proof of Lemma 6

The bad transcript analysis relies on the alternating events lemma proposed by Jha and Nandi [27] and Choi et al. [14].

Lemma 8 (alternating events lemma [14,27]). Let $q_i, q_j, q_k, q_l, q \in \mathbb{N}$. Let $X^{q_i} = (X_1, \ldots, X_{q_i})$ be a q_i -tuple of random variables, and define the same for $X^{q_j}, X^{q_k}, X^{q_l}$. For distinct $i \in [q_i], j \in [q_j]$, let $E_{i,j}$ be events associated with $X_i \in X^{q_i}$ and $X_j \in X^{q_j}$, possibly dependent, which all hold with probability at most ϵ . For distinct $i \in [q_i], j \in [q_j], k \in [q_k], l \in [q_l]$, let $F_{i,j,k,l}$ be events associated with $X_i \in X^{q_i}, X_j \in X^{q_j}, X_k \in X^{q_k}, and X_l \in X^{q_l}$ which all hold with probability at most ϵ' . Moreover, the collection of events $(F_{i,j,k,l})_{i,j,k,l}$ is independent with the collection of event $(E_{i,j})_{i,j}$. Then,

$$\Pr[\exists i \in [q_i], j \in [q_j], k \in [q_k], l \in [q_l], E_{i,j} \wedge E_{k,l} \wedge F_{i,j,k,l}] \le \sqrt{q_i q_j q_k q_l} \cdot \epsilon \cdot \sqrt{\epsilon'}.$$

Jha and Nandi [27] proved the alternating events lemma for $q_i, q_j, q_k, q_l = q$, the lemma can straightforwardly be generalized to different q_i, q_j, q_k, q_l , a similar proof for this is given in the bad transcripts analysis of the work by Choi et al. [14].

We want to bound the probability $\Pr[X_{\mathcal{P}} \in \mathcal{T}_{\text{bad}}]$ that an ideal world transcript τ satisfies one of the bad events. Therefore, by the union bound, the probability that $X_{\mathcal{P}} \in \mathcal{T}_{\text{bad}}$ can be bounded as

$$\Pr[X_{\mathcal{P}} \in \mathcal{T}_{\text{bad}}] \le \sum_{i} \Pr[\text{bad}_{i}].$$

Events bad₁, bad₂ and bad₃. For any possible MAC query $(j, N, M, T) \in \tau_m$, $(u, v) \in \tau_1$ and $(x, y) \in \tau_2$, one has, by the fact that K_j and h_j are independently drawn,

$$\Pr[(K_j = N \oplus u) \land (h_j(M) = N \oplus x)] = \frac{\epsilon}{2^n}.$$

Hence, summing over q_m possible choices of (j, N, M, T), p possible (u, v), and p possible (x, y) yields

$$\Pr[\text{bad}_1] \le \frac{q_m p^2 \epsilon}{2^n} \,.$$

Similarly

$$\Pr[\text{bad}_2] \le \frac{q_m p^2 \epsilon}{2^n} \,,$$

$$\Pr[\text{bad}_3] \le q_m p^2 \epsilon^2 \,.$$

Event bad₄. For this bad event, we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $K_j = K_{j'}$, and the first equation becomes $N \neq N'$. Let \mathcal{N} be the set of all MAC queries such that $N \neq N'$. Event bad₄ occurs if $N' \oplus h_{j'}(M') = N' \oplus$ $h_{j''}(M'')$ for some $(j',N',M',T') \neq (j'',N'',M'',T'')$. For any such fixed $(j,N,M,T),(j',N',M',T'),(j'',N'',M'',T'') \in \tau_m$, one has, by the fact that $h_{j'}$ and $h_{j''}$ are drawn independently, $\Pr[h_{j'}(M') \oplus h_{j''}(M'') = N' \oplus N'] \leq \epsilon$. Hence, summing over at most μ^2 possible (j,N,M,T),(j',N',M',T'),(j'',N'',M'',T''), one obtains

$$\Pr[\operatorname{bad}_4(a)] \le \mu^2 \epsilon$$

2. For case (b), the keys K_j and $K_{j'}$ are generated independently of each other. For any MAC queries $(j, N, M, T), (j', N', M', T'), (j'', N'', M'', T'') \in \tau_m$ with (j', N', M', T') distinct from (j, N, M, T) and from (j'', N'', M'', T''), one has, by the fact that $K_j, K_{j'}, h_{j'}$, and $h_{j''}$ are drawn independently,

$$\Pr[(K_j \oplus K_{j'} = N \oplus N') \land (h_{j'}(M) \oplus h_{j''}(M'') = N \oplus N'')] \le \frac{\epsilon}{2^n}$$

Hence, summing over at most q_m^3 possible (j, N, M, T), (j', N', M', T'), (j'', N'', M'', T''), one obtains

$$\Pr[\operatorname{bad}_4(b)] \le \frac{q_m^3 \epsilon}{2^n}$$

Putting the two cases together, we have

$$\Pr[\mathrm{bad}_4] \le \mu^2 \epsilon + \frac{q_m^3 \epsilon}{2^n}.$$

Events bad₅ and bad₆. We consider event bad₅. For this bad event, we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $K_j = K_{j'}$, and the first equation becomes $N \neq N'$. For any such fixed $(j, N, M, T), (j', N', M', T') \in \tau_m$, the probability of the event $h_j^*(M) \oplus h_{j'}^*(M') = T \oplus T'$ is at most ϵ . Number of queries such that N = N' is at most μ^2 . Summing over all possible (j, N, M, T), (j', N', M', T'), one obtains

$$\Pr[\operatorname{bad}_5(a)] \le \mu^2 \epsilon$$
.

2. For case (b), the keys K_j and $K_{j'}$ are generated independently of each other. For any two distinct MAC queries $(j, N, M, T) \neq (j', N', M', T') \in \tau_m$, one has, by the fact that K_j , $K_{j'}$, h_j^* , and $h_{j'}$ are drawn independently,

$$\Pr[(K_j \oplus K_{j'} = N \oplus N') \land (h_j^*(M) \oplus h_{j'}^*(M') = T \oplus T')] \le \frac{\epsilon}{2^n}$$

Hence, summing over the $q_m(q_m - 1)/2$ possible pairs of $(j, N, M, T) \neq (j', N', M', T')$, we get

$$\Pr[\operatorname{bad}_5(b)] \le \frac{q_m^2 \epsilon}{2^{n-1}} \le \frac{q_m^3 \epsilon}{2^{n-1}}$$

Putting the two cases together, we have

$$\Pr[\mathrm{bad}_5] \le 2\mu\epsilon + \frac{q_m^3\epsilon}{2^{n-1}} \,.$$

Similar to $bad_5(b)$,

$$\Pr[\mathrm{bad}_6] \le \frac{q_m^2 \epsilon^2}{2} \le \frac{q_m^3 \epsilon^2}{2} \,.$$

Events bad₇ and bad₈. We first consider event bad₇. For any two distinct MAC queries $(j, N, M, T) \neq (j', N', M', T') \in \tau_m$ and any $(u, v) \in \tau_1$, one has, by the fact that K_j , h_j , and $h_{j'}$ are drawn independently,

$$\Pr[(K_j = N \oplus u) \land (h_j(M) \oplus h_{j'}(M') = N \oplus N')] \le \frac{\epsilon}{2^n}.$$

Then, summing over $q_m(q_m-1)/2$ possible pairs of $(j, N, M, T) \neq (j', N', M', T')$, and p possible (u, v),

$$\Pr[\text{bad}_7] \le \frac{q_m^2 p \epsilon}{2^{n-1}} \,.$$

We now consider event bad₈. For this bad event, we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $K_j = K_{j'}$, and the second equation becomes $N \neq N'$. For any two distinct MAC queries $(j, N, M, T) \neq (j', N', M', T') \in \tau_m$ and any $(u, v) \in \tau_1$, one has, by the fact that h_j is drawn independently, the event $N \oplus h_j(M) = u$ holds with probability at most ϵ . The number of choices of possible N = N' is at most 2μ and summing over p possible (u, v),

$$\Pr[\operatorname{bad}_8(a)] \le 2\mu p\epsilon$$
.

2. For case (b), the keys K_j and $K_{j'}$ are generated independently of each other. For any two distinct MAC queries $(j, N, M, T) \neq (j', N', M', T') \in \tau_m$, and any $(u, v) \in \tau_1$, one has, by the fact that K_j , $K_{j'}$, and h_j are drawn independently,

$$\Pr[(h_j(M) = N \oplus x) \land (K_j \oplus K_{j'} = N \oplus N')] \le \frac{\epsilon}{2^n}.$$

Then, summing over $q_m(q_m-1)/2$ possible pairs of $(j, N, M, T) \neq (j', N', M', T')$, and p possible (x, y),

$$\Pr[\operatorname{bad}_8(b)] \le \frac{q_m^2 p \epsilon}{2^{n-1}} \,.$$

Putting the two cases together, we have

$$\Pr[\text{bad}_8] \le 2\mu p\epsilon + \frac{q_m^2 p\epsilon}{2^{n-1}}.$$

Events bad_9 and bad_{10} . In order to bound these two events, we will need the following two help events

$$\begin{aligned} \operatorname{bad}_{c_1} &: c_1 = |(j, T, M, C) \in \tau_m \colon N \oplus K_j \in \tau_1| \ge \sqrt{q_m} \,, \\ \operatorname{bad}_{c_2} &: c_2 = |(j, T, M, C) \in \tau_m \colon N \oplus h_j(M) \in \tau_2| \ge \sqrt{q_m} \,. \end{aligned}$$

In stead of bounding bad₉, we will bound $bad_9 \lor bad_{c_1}$

 $\Pr[\operatorname{bad}_9 \lor \operatorname{bad}_{c_1}] = \Pr[\operatorname{bad}_{c_1}] + \Pr[\operatorname{bad}_9 \land \neg \operatorname{bad}_{v_1}].$

To upper bound $\Pr[\text{bad}_{c_1}]$, we see c_1 as a random variable over the random choice of K_i . First, note that

$$\boldsymbol{E}[c_1] = \sum_{(j,N,M,T)\in\tau_m} \sum_{(u,v)\in\tau_1} \Pr[N \oplus K_j = u] = \frac{q_m p}{2^n},$$

so that by Markov's inequality,

$$\Pr[\operatorname{bad}_{c_1}] \le \frac{\sqrt{q_m}p}{2^n}$$

For this bad event $\operatorname{bad}_9 \wedge \neg \operatorname{bad}_{c_1}$, we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $K_j = K_{j'}$. Note that $c_1 < \sqrt{q_m}$, and fix any queries $(j, N, M, T) \neq (j', N', M', T') \in \tau_m, (u, v), (u', v') \in \tau_1$ such that $N \oplus K_j = u$ and $N' \oplus K_j = u'$. Note that since $c_1 < \sqrt{q_m}$, there are at most $q_m/2$ such tuple of queries. Then

$$\Pr[h_i^*(M) \oplus h_i^*(M') = v \oplus T \oplus v' \oplus T'] \le \epsilon.$$

and, by summing over every h_j^* such that $c_1 < \sqrt{q_m}$ and every such tuple of queries, one has

$$\Pr[\operatorname{bad}_9(a) \land \neg \operatorname{bad}_{c_1}] \le \frac{q_m \epsilon}{2} \le \frac{q_m^3 \epsilon^2}{2}.$$

2. For case (b), the keys K_j and $K_{j'}$ are generated independently of each other. Using the fact that K_j , $K_{j'}$, h_j^* , and $h_{j'}^*$ are drawn independently, then, summing over at most q_m^2 possible (j, N, M, T), (j', N', M', T'), and p^2 possible (u, v), (u', v'), one obtains

$$\Pr[\operatorname{bad}_9(b) \lor \operatorname{bad}_{c_1}] = \Pr[\operatorname{bad}_9(b)] \le \frac{q_m^2 p^2 \epsilon}{2^{2n}} \le \frac{q_m p^2 \epsilon}{2^n} \,.$$

Finally,

$$\Pr[\operatorname{bad}_9 \vee \operatorname{bad}_{c_1}] \le \frac{\sqrt{q_m}p}{2^n} + \frac{q_m^3 \epsilon^2}{2} + \frac{q_m p^2 \epsilon}{2^n}.$$

Similarly,

$$\Pr[\operatorname{bad}_{10} \lor \operatorname{bad}_{c_2}] \le \frac{\sqrt{q_m}p}{2^n} + \frac{q_m^3\epsilon^2}{2} + q_m p^2\epsilon^2.$$

Events bad₁₁. Note that since the hash keys for different users are generated independently, it is enough to focus on the single user when $h_{j_{i_1}} = \cdots = h_{j_{i_{\xi+1}}}$, as this case will yield the dominant term. Since $(N^{(j_i)}, M^{(j_i)}) \neq (N^{(j_k)}, M^{(j_k)})$ for all $i \neq k$, we can apply Lemma 2 with $a = \xi = 2^n/2(p+q_m)$ to get

$$\Pr[\text{bad}_{11}] \le \frac{8q_m^2(p+q_m)^2\epsilon}{2^{2n}}.$$

Event bad₁₂. For any MAC query $(j, N, M, T) \in \tau_m$ and any verification query $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)}) \in \tau_a$, one has, by the fact that h_j , $h_{j^{(a)}}$, h_j^* , and $h_{j^{(a)}}^*$ are drawn independently,

$$\Pr[(h_j(M) \oplus h_{j^{(a)}}(M^{(a)}) = N \oplus N^{(a)}) \land (h_j^*(M) \oplus h_{j^{(a)}}^*(M^{(a)}) = T \oplus T^{(a)})] \le \epsilon^2.$$

Then, summing over q_m possible (j, N, M, T), and q_a possible $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)})$,

$$\Pr[\text{bad}_{12}] \le q_m q_a \epsilon^2 \le q_m^2 q_a \epsilon^2$$

Event bad₁₃. For any possible verification query $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)}) \in \tau_a$ and a pair of any possible primitive queries $(u, v) \in \tau_1$ and $(x, y) \in \tau_2$, one has, by the fact that $K_{j^{(a)}}$ and $h_{j^{(a)}}$ are drawn independently,

$$\Pr[(K_{j^{(a)}} = N^{(a)} \oplus u) \land (h_{j^{(a)}}(M^{(a)}) = N^{(a)} \oplus x)] \le \frac{\epsilon}{2^n}.$$

Then, summing over q_a possible $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)})$, p possible (u, v), and p possible (x, y),

$$\Pr[\text{bad}_{13}] \le \frac{q_a p^2 \epsilon}{2^n}.$$

Events bad_{14} and bad_{16} . We consider event bad_{14} . Note that since our goal is the prove 2n/3-bit security, it is sufficient the only focus on the first two equations. For any MAC query $(j, N, M, T) \in \tau_m$, any verification query $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)}) \in \tau_a$, and any $(u, v) \in \tau_1$, one has, by the fact that K_j , h_j , and $h_{j^{(a)}}$ are drawn independently,

$$\Pr[(K_j = N \oplus u) \land (h_j(M) \oplus h_{j^{(a)}}(M^{(a)}) = N \oplus N^{(a)})] \le \frac{\epsilon}{2^n}.$$

Then, summing over q_m possible (j, N, M, T), q_a possible $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)})$, and p possible (x, y),

$$\Pr[\text{bad}_{14}] \le \frac{pq_m q_a \epsilon}{2^n}.$$

Similarly,

$$\Pr[\operatorname{bad}_{16}] \le \frac{pq_m q_a \epsilon}{2^n}.$$

Event bad₁₅. For this bad event, we need to consider the following cases, namely when (a) j = j' (the two queries are made to the same user oracle), and when (b) $j \neq j'$ (the two queries are made to the different user oracles).

1. For case (a), since the *j*-th and the *j'*-th user are the same, we have $K_j = K_{j^{(a)}}$. We will use Lemma 8 to bound the event, with $q_i = q_m$, $q_k = q_a$, $q_j = q_l = p$. We denote $E_{i,j} : N \oplus x = H_j$, $E_{k,\ell} : N^{(a)} \oplus x' = H_{j^{(a)}}$, and $F_{i,j,k,\ell} : T^{(a)} \oplus H^*_{j^{(a)}} = y \oplus T \oplus H^*_j \oplus y'$. The probability that $E_{i,j}$ happens for fixed i, j is ϵ (same for $E_{k,\ell}$), and the probability that $F_{i,j,k,\ell}$ happens for fixed j, k is ϵ . Hence, summing over at most q_m possible (j, N, M, T), q_a possible $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)})$, and p^2 possible (x, y), (x', y'), one obtains

$$\Pr[\operatorname{bad}_{15}(a)] \le p\sqrt{q_m q_a} \epsilon^{3/2}$$

2. For case (b), the keys K_j and $K_{j^{(a)}}$ are generated independently of each other. It is sufficient the only focus on the first two equations. For any MAC queries $(j, N, M, T) \in \tau_m$, any verification queries $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)}) \in \tau_a$, and any $(x, y) \in \tau_2$, one has, by the fact that K_j , $K_{j^{(a)}}$, and h_j are drawn independently,

$$\Pr[(h_j(M) = N \oplus x) \land (K_j \oplus K_{j^{(a)}} = N \oplus N^{(a)})] \le \frac{\epsilon}{2^n}.$$

Hence, summing over at most q_m possible (j, N, M, T), q_a possible $(j^{(a)}, N^{(a)}, M^{(a)}, T^{(a)})$, and p possible (x, y), one obtains

$$\Pr[\operatorname{bad}_{15}(b)] \le \frac{pq_m q_a \epsilon}{2^n} \,.$$

Putting the two cases together, we have

$$\Pr[\text{bad}_{15}] \le p\sqrt{q_m q_a} \epsilon^{3/2} + \frac{p q_m q_a \epsilon}{2^n} \,.$$

The result follows by an union bound over all conditions.