Encrypt or Decrypt? To Make a Single-Key Beyond Birthday Secure Nonce-Based MAC *

Nilanjan Datta[1], Avijit Dutta[2], Mridul Nandi[2] and Kan Yasuda[3]

[1] Indian Institute of Technology, Kharagpur
[2] Indian Statistical Institute, Kolkata
[3] NTT Secure Platform Laboratories, NTT Corporation, Japan.
nilanjan_isi_jrf@yahoo.com, avirocks.dutta13@gmail.com,
mridul.nandi@gmail.com, yasuda.kan@lab.ntt.co.jp

Abstract. In CRYPTO 2016, Cogliati and Seurin have proposed a highly secure nonce-based MAC called Encrypted Wegman-Carter with Davies-Meyer (EWCDM) construction, as $\mathsf{E}_{K_2}\big(\mathsf{E}_{K_1}(N) \oplus N \oplus \mathsf{H}_{K_h}(M)\big)$ for a nonce N and a message M. This construction achieves roughly $2^{2n/3}$ bit MAC security with the assumption that E is a PRP secure n-bit block cipher and H is an almost xor universal n-bit hash function. In this paper we propose Decrypted Wegman-Carter with Davies-Meyer (DWCDM) construction, which is structurally very similar to its predecessor EWCDM except that the outer encryption call is replaced by decryption. The biggest advantage of DWCDM is that we can make a truly single key MAC: the two block cipher calls can use the same block cipher key $K = K_1 = K_2$. Moreover, we can derive the hash key as $K_h = \mathsf{E}_K(1)$, as long as $|K_h| = n$. Whether we use encryption or decryption in the outer layer makes a huge difference; using the decryption instead enables us to apply an extended version of the mirror theory by Patarin to the security analysis of the construction. DWCDM is secure beyond the birthday bound, roughly up to $2^{2n/3}$ MAC queries and 2^n verification queries against nonce-respecting adversaries. DWCDM remains secure up to $2^{n/2}$ MAC queries and 2^n verification queries against nonce-misusing adversaries.

Keywords: EDM, EWCDM, Mirror Theory, Extended Mirror Theory, H-Coefficient.

1 Introduction

Almost all symmetric-key cryptographic systems that use secret keys, including encryption, authentication and authenticated-encryption, are realized via Pseudo-Random Functions (PRFs). Therefore, the study of practical candidates of PRF is essential in providing solutions for the increasing use of cryptography in real-world applications. But unfortunately, very few PRFs are actually available in practice, and it is not easy to construct a sufficiently secure PRF.

* This is the full version of the article accepted in IACR-CRYPTO 2018.

As a result, Pseudo-Random Permutations (PRPs) or block ciphers, which are available in plenty [17, 21, 12, 11], replace the PRF and are deployed as building blocks for almost every cryptographic systems.

Although various block ciphers [17, 21, 12, 11] are available and can be assumed to be PRFs, such an assumption comes at the cost of security degradation due to the PRF-PRP switch [6] which tells that a PRF can be replaced by a PRP up to quadratic degradation in security (often called "birthday bound security"). This loss of security is sometimes acceptable in practice if the block size of the cipher is large enough (e.g. AES-128). But with lightweight block ciphers with relatively small block sizes (e.g. 64-bit), whose number has grown tremendously in recent years (e.g. [11, 21, 12, 2, 1]), the quadratic security loss severly limits their applicability, and as a result it seems to be challenging to use these small ciphers in modern-day lightweight cryptography (e.g. Smart Card, RFID etc.).

In order to save these ciphers from obsolescence, various PRP-to-PRF constructions have been proposed in recent years that guarantee higher security than the usual birthday bound security. Such constructions are often called BBB (Beyond Birthday Bound)—i.e., security against more than $2^{n/2}$ queries where n is the block size of the underlying cipher. A popular BBB construction is the XOR of permutations [7, 24, 3, 29].

XOR OF PERMUTATIONS. Bellare et al. [7] suggested a way to construct a PRF from PRPs by taking the xor (more generally sum) of two independent PRPs.

$$\mathsf{XOR}_{\mathsf{E}_{K_1},\mathsf{E}_{K_2}}(x) = \mathsf{E}_{K_1}(x) \oplus \mathsf{E}_{K_2}(x).$$

The construction was later analyzed by Lucks [24] and proved its security up to $2^{2n/3}$ queries, where n is the block size of the underlying PRP. Single-keyed variant of this construction was first analyzed by Bellare and Impagliazzo [3] and showed its security upto $O(nq/2^n)$ by using some advanced probability theory. However, their proof was sketchy and hard to verify. Subsequently, a lot of efforts have been invested towards improving the bound of XOR construction and its single-keyed variant (even proving upto n-bit security) by Patarin as evident in [29, 33, 32], but the proof contains serious gaps. Later in FSE 2014, the result was generalized to the xor of three or more independent PRPs [14]. However, a verifiable n-bit security proof of XOR construction is later provided by Dai et al. [18] using *chi-squared* method. Although, the original proof contained a glitch, which was pointed out by Bhattacharya and Nandi [9], was later fixed in the full version of [18].

The XOR construction provides a solution for encryption by combining itself with the counter (CTR) mode of encryption, as instantiated in CENC construction, a highly efficient nonce-based encryption mode proposed by Iwata [22]. It was shown [22] that CENC, as an encryption mode, achieves security upto $O(2^{2n/3})$ queries against all nonce-respecting adversaries. Later, Iwata et al. [23] provided the optimal security bound of the construction based on mirror theory technique [33]. Recently, Bhattacharya and Nandi [10] have given an optimal se-

curity bound of CENC by analysing the PRF security bound of variable output length of xor of permutations using chi-squared method.

Though useful for encryption, the XOR construction does not seem to be directly usable for authentication. A problem for authentication is that we have to extend the domain size, so that the construction can authenticate long messages. This can be done by hashing the message, but with the XOR construction it seems that we need some subtle combination with a double-block hash function, as employed in PMAC_Plus [34], 1K-PMAC_Plus [19] and LightMAC_Plus [27].

Encrypted Davies-Meyer. The above problem with the XOR construction in authentication was solved by Cogliati and Seurin [15], who proposed a PRP-to-PRF conversion method, called Encrypted Davies-Meyer (EDM). The EDM construction is defined as follows:

$$\mathsf{EDM}_{\mathsf{E}_{K_1},\mathsf{E}_{K_2}}(x) = \mathsf{E}_{K_2}(\mathsf{E}_{K_1}(x) \oplus x).$$

EDM uses two independent block-cipher keys and achieves $O(q^3/2^{2n})$ security [15]. Soon after, Dai et al. [18] improved its bound to $O(q^4/2^{3n})$ by applying chi-squared method. Concurrently, Mennink and Neves [25] proved its almost optimal security, i.e. $O(2^n/67n)$, using mirror theory technique.

As EDM requires two indepenent keys for the cipher calls, it would be interesting to see that whether BBB security holds for the single key setting of the construction, as mentioned by authors of [15]. Recently, Cogliati and Seurin answers the question in positive and proved its security upto $O(q/2^{2n/3})$ [16], as originally conjectured by themselves [15].

Encrypted Wegman-Carter with Davies-Meyer. Following the construction of EDM, Cogliati and Seurin extended the idea to construct EWCDM, a nonce-based BBB secure MAC, which is defined as follows:

$$\mathsf{EWCDM}_{\mathsf{E}_{K_1},\mathsf{E}_{K_2},\mathsf{H}_{K_h}}(N, M) = \mathsf{E}_{K_2}\big(\mathsf{E}_{K_1}(N) \oplus N \oplus \mathsf{H}_K(M)\big)$$

where N is the nonce and M is the message to be authenticated. Note that, EWCDM uses two independent block-cipher keys, K_1 and K_2, and also another independent hash-key K_h for the AXU hash function. [4] In this way, EDM obviated the necessity of using double-block hash function that existed with the XOR construction. It has been proved that EWCDM is secure against all nonce-respecting MAC adversaries[5] that make at most $2^{2n/3}$ MAC queries and 2^n verification queries. Cogliati and Seurin also proved $O(2^{n/2})$ security of the construction against nonce-misusing adversaries. Later, Mennink and Neves [25] proved its n-bit PRF security using mirror theory in the nonce respecting setting and mentioned that the analysis straightforwardly generalizes to the analysis for unforgeability or for the nonce-misusing setting of the construction. The trick

[4] An AXU hash function is a keyed hash function such that for any two distinct messages, the probability, over a random draw of a hash key, of the hash differential being equal to a specific output is small.

[5] adversaries who never repeat the same value of N in their MAC queries

involved in proving the optimal security of EWCDM is by replacing the last block cipher call with its inverse. This subtle change does not make any difference in the output distribution and as a bonus, it trivially allows one to express the output of the construction as a sum of two random permutations (or in general a bi-variate affine equation [6]). It is only this feature which is captured by the *mirror theory* to derive the security bound of the construction.

Motivation behind This Work. Constructions with multiple secret keys necessarily requires larger storage space for storing them, which is sometimes infeasible for lightweight crypto devices like RFID, Smart Card etc. All popular MACs, including CMAC [28] and HMAC [4], require only a single secret key. But most of the time reducing the number of keys without compromising the security is not a trivial task.

As evident from the definition of the construction, EWCDM requires three keys; two block cipher keys K_1 and K_2 and one hash key K_h. Thus, it is natural to ask that whether one can achieve the similar security in the case of using less number of keys. Cogliati and Seurin [15] believed that BBB security should hold for single-keyed EWCDM (with $K_1 = K_2$) but be likely cumbersome to prove. In fact, proving BBB security of single-keyed EDM, as mentioned earlier, is a highly complicated task as evident from [16] and it is not clear at all how to build on this result to prove the MAC security of EWCDM construction with $K_1 = K_2$. Moreover, Cogliati and Seurin, in their proof of single-keyed EDM [16], have also stated that

"For now, we have been unable to extend the current (already cumbersome) counting used for the proof of the single-permutation EDM construction to the more complicated case of single-key EWCDM."

Thus, we expect that proving the MAC security of single-keyed EWCDM should be a notably hard task and very likely require heavy mathematical tools like *Sum Capture Lemma* as already used for single-keyed EDM. This motivates us to design an another single-keyed, nonce-based MAC built from block ciphers (and a hash function) with BBB security that can be proven by a simpler approach.

Our Contribution. Our contribution in this paper is fourfold which we outline as follows:

- DWCDM: New Nonce-Based MAC. We propose *Decrypted Wegman-Carter with Davies-Meyer*, in short DWCDM, a nonce-based BBB secure MAC. The design philosophy of DWCDM is inspired from the trick used in [25] while proving the optimal security of EWCDM. Recall that, in [25], authors replace the last block cipher call with its inverse so that the output of EWCDM can be expressed as a sum of *two independent* PRPs. But the same trick does not work at the time of using the same block cipher key in the construction.

[6] For two variables, P, Q and $\lambda \in \mathrm{GF}(2^n)$ we call an equation of the form $P \oplus Q = \lambda$, a bi-variate affine equation

This phenomenon triggers us to design a nonce based MAC, very similar to EWCDM, in which instead of using the encryption algorithm in the last block-cipher call, we use its **decryption** algorithm so that the output of the construction can be expressed as a sum of two *identical* PRPs and hence the name **Decrypted** Wegman-Carter with Davies-Meyer. The construction is *single-keyed* in the sense that the *same block cipher key* is used for the two cipher calls. Schematic diagram of DWCDM is shown in Fig.4.1 where the last $n/3$ bits of the nonce N is zero, i.e. $N = N^*\|0^{n/3}$. We would like to mention here that one cannot use the full n-bit nonce in DWCDM as that would end up with a birthday bound MAC attack which is described in section 4.1. We show that DWCDM is secure up to $2^{2n/3}$ MAC queries and 2^n verification queries against nonce-respecting adversaries. We also show that DWCDM is secure up to $2^{n/2}$ MAC queries and 2^n verification queries in the nonce-misuse setting, where the bound is tight. As a concrete example of DWCDM, we present an instantiation of DWCDM with the AXU hash function being realized via PolyHash [26]. We show that nPolyMAC achieves $2^{2n/3}$-bit MAC security in the nonce-respecting setting.

- Extended Mirror Theory. Since, our study of interest is the MAC security of the construction, we require to analyze the number of solutions of a system of affine bi-variate equations along with *affine uni-variate and bi-variate non-equations* [7]. Such a general treatment of analysing system of affine equations with non-equations was only mentioned in [33] without giving any formal analysis. To the best of our knowledge, this is the first time we analyse such a generic system of equations with non-equations, which we regard to as *extended mirror theory* and our MAC security proofs of DWCDM and 1K-DWCDM are crucially based on this new result.
- 1K-DWCDM: "Pure" Single-Keyed Variant of DWCDM. Moreover, we exhibit a truly single-keyed nonce-based MAC construction, 1K-DWCDM. Under the condition that the length of the hash key is equal to the block size as $|K_h| = n$, we can even derive the hash key as $K_h = \mathsf{E}_K(0^{n-1}\|1)$, which results in the construction 1K-DWCDM. We prove that 1K-DWCDM is essentially as secure as DWCDM.
- Potentiality of Achieving Higher Security. Finally, we show how one can boost the security for DWCDM type constructions using extended generalized version of Mirror Theory.

Proof Approach. Our MAC security proof of DWCDM and 1K-DWCDM is fundamentally relied on Patarin's H-coefficient technique [30]. Similar to the technique of [15, 20], we cast the unforgeability game of MAC to an equivalent indistinguishablity game, with some suitable choice of ideal world, that allows us to apply the H-coefficient technique for bounding the distinguishing advantage of the construction of our concern.

[7] For two variables, P, Q and $\lambda \in \mathrm{GF}(2^n) \setminus 0^n$ we call $P \oplus Q \neq \lambda$, an affine bi-variate non-equation and $P \neq \lambda$ is an affine uni-variate non-equation.

As mentioned earlier that one can express the output of DWCDM as a sum of two identical permutations. Thus, q many such evaluations of DWCDM gives us a system of q many affine bi-variate equations

$$\begin{cases} \mathsf{E}_K(N_1) \oplus \mathsf{E}_K(T_1) = N_1 \oplus H_{K_h}(M_1) \\ \mathsf{E}_K(N_2) \oplus \mathsf{E}_K(T_2) = N_2 \oplus H_{K_h}(M_2) \\ \qquad\qquad\qquad\qquad \vdots \\ \mathsf{E}_K(N_q) \oplus \mathsf{E}_K(T_q) = N_q \oplus H_{K_h}(M_q) \end{cases}$$

Along with this, we also need to ensure that the verification attempt of the adversary should fail (as a part of the good transcript), i.e. for a verification query (N', M', T'), chosen by the adversary, we should always have

$$\mathsf{E}_K^{-1}(\mathsf{E}_K(N') \oplus N' \oplus H_{K_h}(M')) \neq T'.$$

Hence, it tells us that we also need to incorporate *affine non-equations* along with the system of bi-variate affine equations. This leads us to *extend* the mirror theory technique (extension as in incorporating affine non-equations along with affine bi-variate equations). We require the result of extended mirror theory while lower bounding the real interpolation probability for a good transcript.

Remark 1. Recently, a new technique called *chi-square method* has been evolved as a tool for upper bounding the statistical distance between the answers of two interactive systems and is typically used to prove the information-theoretic pseudorandomness of constructions. This method has been successfully applied in proving the security of XOR construction [18], EDM construction [18], swap or not construction [18] and variable output length XOR construction [9]. Recently, Bhattacharya and Nandi [8] have shown the full indifferential security of xor of two or more random permutations using this method. Moreover, it has been shown in [18] that for some setting, chi-square method outperforms H-coefficient technique in terms of guarranting security with quadratic improvement on the number of queries that adversary can make. However, in our construction, we find it difficult to apply this method as chi-square requires high entropy of the conditional distribution. In D(or E)WCDM, one needs to condition the hash key. Unlike random permutation, the hash key doesnt have sufficient entropy in the conditional probability and there lies the difficulty of applying this method. This fact also acts as a motivation for us to consider DWCDM so that we can represent the construction as a sum of permutations and eventually apply extended mirror theory.

2 Preliminaries

Symbols and Notations. For a set $\mathcal{X}$, $X \leftarrow_{\$} \mathcal{X}$ denotes that X is sampled uniformly at random from $\mathcal{X}$ and independent to all random variables defined so far. $\{0,1\}^n$ denotes the set of all binary strings of length n. The set of all

functions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted as $\mathsf{Func}(\mathcal{X},\mathcal{Y})$ and the set of all permutations over $\mathcal{X}$ is denoted as $\mathsf{Perm}(\mathcal{X})$. $\mathsf{Func}_{\mathcal{X}}$ denotes the set of all functions from $\mathcal{X}$ to $\{0,1\}^n$ and Perm denotes the set of all permutations over $\{0,1\}^n$. We often write Func instead of $\mathsf{Func}_{\mathcal{X}}$ when the domain of the functions is understood from the context. We write $[q]$ to refer to the set $\{1,\ldots,q\}$.

For any binary string x, $|x|$ denotes the length i.e. the number of bits in x. For $x,y \in \{0,1\}^n$, we write $z = x \oplus y$ to denote the modulo 2 addition of x and y. We write $\mathbf{0}$ to denote the zero element of the field $\{0,1\}^n$ (i.e. 0^n) and $\mathbf{1}$ to denote $0^{n-1}\|1$. For integers $1 \le b \le a$, we write $(a)_b$ to denote $a(a-1)\ldots(a-b+1)$, where $(a)_0 = 1$ by convention.

2.1 Security Definitions

PRF AND PRP AND SPRP. A keyed function with key space $\mathcal{K}$, domain $\mathcal{X}$ and range $\mathcal{Y}$ is a function $\mathsf{F} : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ and we denote $\mathsf{F}(K,X)$ by $\mathsf{F}_K(X)$. Similarly, a keyed permutation with key space $\mathcal{K}$ and domain $\mathcal{X}$ is a mapping $\mathsf{E} : \mathcal{K} \times \mathcal{X} \to \mathcal{X}$ such that for all key $K \in \mathcal{K}$, $X \mapsto \mathsf{E}(K,X)$ is a permutation over $\mathcal{X}$ and we denote $\mathsf{E}_K(X)$ for $\mathsf{E}(K,X)$.

PRF. Given an oracle algorithm A with oracle access to a function from $\mathcal{X}$ to $\mathcal{Y}$, making at most q queries, running time is at most t and outputting a single bit. We define prf-advantage of A against the family of keyed functions F as

$$\mathbf{Adv}^{\mathrm{PRF}}_{\mathsf{F}}(\mathsf{A}) := |\Pr[K \leftarrow_{\$} \mathcal{K} : \mathsf{A}^{\mathsf{F}_K} = 1] - \Pr[\mathsf{RF} \leftarrow_{\$} \mathsf{Func}(\mathcal{X},\mathcal{Y}) : \mathsf{A}^{\mathsf{RF}} = 1]|.$$

We say that F is (q,t,ϵ) secure PRF, if $\mathbf{Adv}^{\mathrm{PRF}}_{\mathsf{F}}(q,t) := \max_{\mathsf{A}} \mathbf{Adv}^{\mathrm{PRF}}_{\mathsf{F}}(\mathsf{A}) \le \epsilon$, where the maximum is taken over all adversaries A that makes q many queries and running time is at most t.

PRP. Given an oracle algorithm A with oracle access to a permutation of $\mathcal{X}$, making at most q queries, running time is at most t and outputting a single bit. We define prp-advantage of A against the family of keyed permutations E as

$$\mathbf{Adv}^{\mathrm{PRP}}_{\mathsf{E}}(\mathsf{A}) := |\Pr[K \leftarrow_{\$} \mathcal{K} : \mathsf{A}^{\mathsf{E}_K} = 1] - \Pr[\Pi \leftarrow_{\$} \mathsf{Perm}(\mathcal{X}) : \mathsf{A}^{\Pi} = 1]|.$$

We say that E is (q,t,ϵ) secure PRP, if $\mathbf{Adv}^{\mathrm{PRP}}_{\mathsf{E}}(q,t) := \max_{\mathsf{A}} \mathbf{Adv}^{\mathrm{PRP}}_{\mathsf{E}}(\mathsf{A}) \le \epsilon$, where the maximum is taken over all adversaries A that makes q many queries and running time is at most t.

SPRP. Given an oracle algorithm A with oracle access to a permutation and its inverse over $\mathcal{X}$, making at most q^+ queries to permutation and q^- queries to inverse permutation, running time is at most t and outputting a single bit. We define sprp-advantage of A against the family of keyed permutations E as

$$\mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(\mathsf{A}) := |\Pr[K \leftarrow_{\$} \mathcal{K} : \mathsf{A}^{\mathsf{E}_K,\mathsf{E}_K^{-1}} = 1] - \Pr[\Pi \leftarrow_{\$} \mathsf{Perm}(\mathcal{X}) : \mathsf{A}^{\Pi,\Pi^{-1}} = 1]|.$$

We say that E is (q,t,ϵ) secure SPRP, if $\mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q,t) := \max_{\mathsf{A}} \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(\mathsf{A}) \le \epsilon$, where the maximum is taken over all adversaries A that makes q many encryption and decryption queries altogether and running time is at most t.

MACs. Given four non-empty finite sets $\mathcal{K}, \mathcal{N}, \mathcal{M}$ and $\mathcal{T}$, a nonce based keyed function with key space $\mathcal{K}$, nonce space $\mathcal{N}$, message space $\mathcal{M}$ and range $\mathcal{T}$ is a keyed function whose domain is $\mathcal{N} \times \mathcal{M}$ and range is $\mathcal{T}$ and we write $\mathsf{F}(K, N, M)$ as $\mathsf{F}_K(N, M)$.

Definition 1 (Nonce Based MAC). *Let $\mathcal{K}, \mathcal{N}, \mathcal{M}$ and $\mathcal{T}$ be four non-empty finite sets and $\mathsf{F} : \mathcal{K} \times \mathcal{N} \times \mathcal{M} \to \mathcal{T}$ be a nonce based keyed function. For $K \in \mathcal{K}$, let Ver_K be the verification oracle that takes as input $(N, M, T) \in \mathcal{N} \times \mathcal{M} \times \mathcal{T}$ and outputs 1 if $\mathsf{F}_K(N, M) = T$, otherwise outputs 0. A (q_m, q_v, t) adversary against the MAC security of F is an adversary A with access to two oracles F_K and Ver_K for $K \in \mathcal{K}$ such that it makes at most q_m many MAC queries to first oracle and q_v many verification queries to second oracle. We say that A forges F if any of its queries to Ver_K returns 1. The advantage of A against the MAC security of F is defined as*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{F}}(\mathsf{A}) := \Pr[K \leftarrow\!\!\$\, \mathcal{K} : \mathsf{A}^{\mathsf{F}_K, \mathsf{Ver}_K} \text{ forges }],$$

where the probability is taken over the randomness of the underlying key and the random coin of adversary A (if any). We assume that A does not make any verification query (N, M, T) to Ver_K if T is obtained in previous MAC query with input (N, M) and it does not repeat any query. We call such an adversary as "non-trivial" adversary. The adversary is said to be "nonce respecting" if it does not repeat nonces in its queries to the MAC oracle [8].

r-way Regular And AXU Hash Function. We will need the following definition of a r-way regular and an almost xor-universal (AXU) hash function.

Definition 2. *Let $\mathcal{K}_h, \mathcal{X}, \mathcal{Y}$ be three non-empty finite sets and H be a keyed function $\mathsf{H} : \mathcal{K}_h \times \mathcal{X} \to \mathcal{Y}$. Then,*

1. *H is said to be an ϵ-r-way regular hash function if for any distinct $X_1, X_2, \ldots, X_r \in \mathcal{X}$ and for any non-zero $Y \in \mathcal{Y}$,*

$$\Pr[K_h \leftarrow\!\!\$\, \mathcal{K}_h : \mathsf{H}_{K_h}(X_1) \oplus \cdots \oplus \mathsf{H}_{K_h}(X_r) = Y] \leq \epsilon. \tag{1}$$

   When $r = 1$, then we will often refer to ϵ-1-way regular hash function as simply ϵ-regular hash function.

2. *H is said to be an ϵ-almost xor universal (AXU) hash function if for any distinct $X, X' \in \mathcal{X}$ and for any $Y \in \mathcal{Y}$,*

$$\Pr[K_h \leftarrow\!\!\$\, \mathcal{K}_h : \mathsf{H}_{K_h}(X) \oplus \mathsf{H}_{K_h}(X') = Y] \leq \epsilon. \tag{2}$$

Note that the definition of 2-way regular hash function is almost similar to the definition of AXU hash function, only difference is that the former one requires the output Y to be non-zero whereas the later one excludes such restriction from Y.

[8] Similar to nonce respecting adversary, we say that an adversary is nonce misusing if the adversary is not restricted to make queries to the MAC oracle with distinct nonces.

In the following, we state that PolyHash [26] is one of the examples of algebraic hash function which is $\ell/2^n$-3-way and 1-way regular. Moreover, it is also $\ell/2^n$-AXU hash function.

Proposition 1. *Let* $\mathsf{Poly} : \{0,1\}^n \times (\{0,1\}^n)^\ell \to \{0,1\}^n$ *be a hash function defined as follows: For a fixed key* $K_h \in \{0,1\}^n$ *and a message* $M = (M_1, M_2, \ldots, M_l)$, *where* $|M_i| = n$ *for all* $i = 1, \ldots, l-1$ *and* $|M_l| \leq n$, *we define*

$$\mathsf{Poly}_{K_h}(M) = (M_l \| 0^{n-|M_l|}) \cdot K_h \oplus M_{l-1} \cdot K_h^2 \oplus \ldots \oplus M_1 \cdot K_h^l, \tag{3}$$

where l *is the number of* n*-bit blocks. Then,* Poly *is* $\ell/2^n$*-3-way and* 1*-way regular hash function, where* ℓ *denotes the maximum number of message blocks of size* n*-bits. Moreover it is also* $\ell/2^n$*-AXU hash function.*

Proof. For three distinct messages M, M' and M'', $\mathsf{Poly}_{K_h}(M) \oplus \mathsf{Poly}_{K_h}(M') \oplus \mathsf{Poly}_{K_h}(M'')$ is a polynomial of K_h. Now, if this polynomial reduces to a zero polynomial, which is possible by certain choices of M, M' and M'' (e.g $M = M_1 \| M_2, M' = (M_1 \oplus M_2) \| M_3$ and $M'' = M_2 \| (M_2 \oplus M_3)$), then $\epsilon_3 = 0$. Otherwise, the polynomial is a non-trivial one and hence, $\epsilon_3 = \ell/2^n$, where ℓ is the maximum number of message blocks among three messages. Similarly, it is easy to see that Poly is a $\ell/2^n$-1-way regular hash function as maximum number of roots for the polynomial $\mathsf{Poly}_{K_h}(M) \oplus Y$, where $Y \neq \mathbf{0}$, is ℓ. Moreover, for any two distinct messages M and M', $\mathsf{Poly}_{K_h}(M) \oplus \mathsf{Poly}_{K_h}(M') \oplus Y$ is a non-trivial polynomial of K_h of degree at most ℓ and hence the maximum number of roots this can polynomial can have is ℓ and thus, the AXU advantage becomes $\ell/2^n$. □

3 Patarin's Mirror Theory

Mirror theory, as defined in [33] is the theory of evaluating the number of solutions of affine system of equalities and non-equalities in a finite group. Patarin, who coined this theory, has given a lower bound on the number of solutions of a finite system of affine bi-variate equations using an inductive proof when the variables in the equations are wor samples [31]. The proof is tractable upto the order of $2^{2n/3}$ security bound, but the proof becomes highly complex and too difficult to verify in the case of deriving the optimal security bound. In specific, once the first-order recursion is considered, one needs to consider a second-order recursion, and so on, until the n-th recursion. For the i-th order recursion, there are $O(2^i)$ many cases and Patarin's proof only addresses the first (and perhaps the second) order recursion by a tedious analysis, but the cases of the higher-order ones are quite different, and it's not at all clear how to bridge the gap, given an exponential number of cases that one has to consider. Moreover, to the best of our knowledge, the proof did not consider any affine non-equation as well.

In this section we extend the Mirror theory in the context of our MAC security to incorporate the affine non-equations (that includes uni-variate and

bi-variate non-equations) along with a system of affine bi-variate equations. In the following, we prove that when the number of affine bi-variate equations is $q \leq 2^{2n/3}$ and the number of non-equations is $v \leq 2^n$ (v is the total number of affine uni-variate and bi-variate non equations), then the number of solutions becomes at least $(2^n)_{3q/2}/2^{nq}$. For the sake of presentation and interoperability with the results in the remainder of the paper, we use different parameterization and naming convention.

3.1 General Setting of Mirror Theory

Given a bi-variate affine equation $P \oplus Q = \lambda$, the associated linear equation of this affine equation is $P \oplus Q = 0$. Now, given $\lambda_1, \ldots, \lambda_q \in \mathrm{GF}(2^n) \setminus \mathbf{0}$ which we write as $\Lambda = (\lambda_1, \ldots, \lambda_q)$, let us consider a system of q many bi-variate affine equations over $\mathrm{GF}(2^n)$:

$$\mathcal{E}_\Lambda = \{P_{n_1} \oplus P_{t_1} = \lambda_1, P_{n_2} \oplus P_{t_2} = \lambda_2, \ldots, P_{n_q} \oplus P_{t_q} = \lambda_q\}.$$

Given a function $\phi : \{n_1, t_1, \ldots, n_q, t_q\} \to \mathcal{I}$, called *index mapping function*, we associate another system of bi-variate affine equations:

$$\mathcal{E}_{\Lambda,\phi} = \{P_{\phi(n_1)} \oplus P_{\phi(t_1)} = \lambda_1, P_{\phi(n_2)} \oplus P_{\phi(t_2)} = \lambda_2, \ldots, P_{\phi(n_q)} \oplus P_{\phi(t_q)} = \lambda_q\}.$$

Let α denotes the cardinality of the image set of ϕ. Then, $\mathcal{E}_{\Lambda,\phi}$ is a system of bi-variate affine equations over α variables. In our paper, a specific choice of $\mathcal{I}$ would be $\{0,1\}^n$.

Example. Consider a system of equations:

$$\{P_1 \oplus P_2 = \lambda_1, P_1 \oplus P_3 = \lambda_2, P_2 \oplus P_4 = \lambda_3\}.$$

Then, the index mapping function for the above system of equations is $\phi(n_1) = 1, \phi(t_1) = 2, \phi(n_2) = 1, \phi(t_2) = 3, \phi(n_3) = 2, \phi(t_3) = 4$. For this system of equations $\alpha = 4$.

Equation-Dependent Graph. For index mapping function $\phi : \{n_1, t_1, \ldots, n_q, t_q\} \to \mathcal{I}$, we associate a undirected graph $G_\phi = ([q], \mathcal{S})$ where $\{i, j\} \in \mathcal{S}$ if

$$|\{\phi(n_i), \phi(t_i)\} \cap \{\phi(n_j), \phi(t_j)\}| \geq 1$$

or if $i = j$ and $\phi(n_i) = \phi(t_i)$. We call such an edge a *self-loop*. In other words, we introduce an edge between two equations (node represents the equation number) in the equation-dependent graph if the corresponding equations have at least one common unknown variable. Note that the set $\{\phi(n_i), \phi(t_i)\}$ can be a multi-set.

For a subset $\{i_1, \ldots, i_c\} \subseteq [q]$, let

$$\{P_{\phi(n_{i_1})} \oplus P_{\phi(t_{i_1})} = 0, P_{\phi(n_{i_2})} \oplus P_{\phi(t_{i_2})} = 0, \ldots, P_{\phi(n_{i_c})} \oplus P_{\phi(t_{i_c})} = 0\}$$

be the sub-system of associated linear equations. We say this sub-system of associated linear equations is linearly dependent if $\{i_1, \ldots, i_c\}$ is the minimal set and all variables P_x, which appeared in the above sub-system, appears exactly twice. Depending on the value of c (for the minimal linearly dependent sub-system), we have the following three cases;

(i) $c = 1$: SELF-LOOP. If there exists i such that $\phi(n_i) = \phi(t_i)$.

(ii) $c = 2$: PARALLEL-EDGE. If there exists $i \neq j$ such that either:

(a) $\phi(n_i) = \phi(n_j)$ and $\phi(t_i) = \phi(t_j)$ or (b) $\phi(n_i) = \phi(t_j)$ and $\phi(t_i) = \phi(n_j)$.

(iii) $c \geq 3$: ALTERNATING-CYCLE. If there exists distinct $i_1, i_2, \ldots, i_c$ such that for every $j \in [c]$ either

- $\phi(n_{i_j}) \in \{\phi(n_{i_{j+1}}), \phi(t_{i_{j+1}})\}$ and $\phi(t_{i_j}) \in \{\phi(n_{i_{j-1}}), \phi(t_{i_{j-1}})\}$ or
- $\phi(t_{i_j}) \in \{\phi(n_{i_{j+1}}), \phi(t_{i_{j+1}})\}$ and $\phi(n_{i_j}) \in \{\phi(n_{i_{j-1}}), \phi(t_{i_{j-1}})\}$.

When $i = 1$, $i - 1$ is considered as c and when $i = c$, $i + 1$ is considered as 1. We say that ϕ is *dependent* if any one of the above condition holds. Otherwise, we call it *independent*. Given an independent ϕ, the graph G_ϕ becomes a simple graph and $\mathcal{E}_{\Lambda,\phi}$ becomes linearly independent. In this case, the number of variables present in a connected component $C = \{i_1, \ldots, i_c\}$ of G_ϕ (i.e., the size of the set $\{\phi(n_{i_1}), \phi(t_{i_1}), \ldots, \phi(n_{i_c}), \phi(t_{i_c})\}$) is exactly $c + 1$. We call the set $\{\phi(n_{i_1}), \phi(t_{i_1}), \ldots, \phi(n_{i_c}), \phi(t_{i_c})\}$ a *block*. The *block maximality*, denoted by $\xi_{\max}$, of an independent ϕ is defined as $\zeta_{\max}+1$ where $\zeta_{\max}$ is the size of the maximum connected components of G_ϕ (Note that, a block with p many elements introduces $p - 1$ many affine equations.).

Example. Suppose, $\mathcal{E} = \{(1)\ P_1 \oplus P_2 = \lambda_1, (2)\ P_1 \oplus P_3 = \lambda_2, (3)\ P_2 \oplus P_4 = \lambda_3, (4)\ P_5 \oplus P_6 = \lambda_4\}$, then we have two blocks $\{1, 2, 3, 4\}$ and $\{5, 6\}$ and $\xi_{\max} = 4$. Observe that, we also have two connected components $\{1, 2, 3\}$ and $\{4\}$ and the size of the maximum connected component $\zeta_{\max} = 3$ as depicted in Fig. 3.1 and $\xi_{\max} = \zeta_{\max} + 1 = 4$.

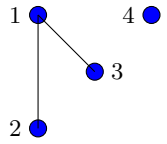

Fig. 3.1. Equation dependent graph $G[\mathcal{E}]$ comprised of two components; largest component size is 3.

3.2 Extended Mirror Theory

In this section, we introduce the extended Mirror theory technique by incorporating two types of non-equations with a finite number of bi-variate affine equations. We consider (i) uni-variate affine non-equation of the form $X_i \neq c$ and (ii) bi-variate affine non-equation of the form $X_i \oplus Y_i \neq c$, where c is a non-zero constant. In particular, we lower bound the number of solutions of a

finite number of affine equations [9] and uni(bi-) variate affine non-equations. To begin with, let us investigate what happens when we introduce a single uni(bi-) variate affine non-equation with a finite number of affine equations.

Let $\mathcal{E}^{=}$ be a system of q many affine equations of the form

$$\mathcal{E}^{=} = \{P_{n_1} \oplus P_{t_1} = \lambda_1, \ldots, P_{n_q} \oplus P_{t_q} = \lambda_q\}. \tag{4}$$

Let ϕ be an index mapping function that maps from $\{n_1, t_1, \ldots, n_q, t_q\} \to \mathcal{I}$. Let $\Lambda_{=} = (\lambda_1, \lambda_2, \ldots, \lambda_q)$, where each $\lambda_i \in \mathrm{GF}(2^n) \setminus \mathbf{0}$. Now, for an independent choice of ϕ, $\mathcal{E}^{=}_{\phi,\Lambda_{=}}$ is a linearly independent set of q many affine equations. Let $\mathcal{E}^{\neq}$ be a system of r many bi-variate affine non-equations and $v-r$ many uni-variate affine non-equations of the form

$$\begin{aligned}\mathcal{E}^{\neq} = \quad & \{P_{n_{q+1}} \oplus P_{t_{q+1}} \neq \lambda'_1, \ldots, P_{n_{(q+r)}} \oplus P_{t_{(q+r)}} \neq \lambda'_r\} \\ & \bigcup \{P_{n_{q+r+1}} \neq \lambda'_{r+1}, \ldots, P_{n_{(q+v)}} \neq \lambda'_v\}.\end{aligned}$$

We denote $\Lambda_{\neq} = (\lambda'_1, \lambda'_2, \ldots, \lambda'_v)$, where each $\lambda'_i \in \mathrm{GF}(2^n) \setminus \mathbf{0}$, and $\Lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_q, \lambda'_1, \lambda'_2, \ldots, \lambda'_v)$. Now, for the system of affine equations and non-equations $\mathcal{E} := \mathcal{E}^{=} \cup \mathcal{E}^{\neq}$, we consider the index mapping function

$$\phi' : \{n_1, t_1, \ldots, n_q, t_q, n_{q+1}, t_{q+1}, \ldots, n_{q+v}, t_{q+v}\} \to \mathcal{I}.$$

Moreover, we denote $\phi := \phi'_{|q}$ to be the index mapping function that maps $\{n_1, t_1, \ldots, n_q, t_q\} \to \mathcal{I}$ and $\Lambda_{=} := \Lambda'_{|q}$ to be $(\lambda_1, \lambda_2, \ldots, \lambda_q)$.

Characterizing Good (ϕ', Λ'). We say that a pair (ϕ', Λ') is good if

- (C1) ϕ is independent and for all $x \neq y$, $P_{\phi(x)} = P_{\phi(y)}$ cannot be generated from the system of equations $\mathcal{E}^{=}_{\phi,\Lambda_{=}}$.
- (C2) for all $j \in [v]$ and $i_1, \ldots, i_c \in [q]$, $c \geq 0$, such that $\{i_1, \ldots, i_c, q+j\}$ is dependent system then $\lambda_{i_1} \oplus \cdots \oplus \lambda_{i_c} \oplus \lambda'_j \neq \mathbf{0}$.

In words, a good (ϕ', Λ') says that: (i) the system of equation $\mathcal{E}_{\phi,\Lambda_{=}}$ is linearly independent system of equations and one cannot generate an equation of the form $P_{\phi(x)} = P_{\phi(y)}$ by linearly combining the equation of $\mathcal{E}_{\phi,\Lambda_{=}}$. Moreover, (ii) by linearly combining the equation of $\mathcal{E}_{\phi,\Lambda_{=}}$, one cannot generate an equation of the form $P_x \oplus P_y = \lambda_{x,y}$ such that $P_x \oplus P_y \neq \lambda_{x,y}$ already exist in $\mathcal{E}^{\neq}_{\phi',\Lambda'}$.

Summarizing above, we state and prove the following main theorem, which we call as *Extended Mirror Theorem for* $\xi_{\max} = 3$. For the notational simplicity we assume the index set $\mathcal{I} = [\alpha]$.

[9] when we consider affine equation, we actually refer to the bi-variate affine equation.

Theorem 1. *Let $(\mathcal{E}^{=} \cup \mathcal{E}^{\neq}, \phi', \Lambda')$ be a system of q many affine equations and v many uni(bi-) variate affine non-equations associated with index mapping function ϕ' over $\mathrm{GF}(2^n)$ which are of the form*

$$\begin{aligned}
&(a) P_{\phi(n_i)} \oplus P_{\phi(t_i)} = \lambda_i (\neq \mathbf{0}),\ \forall i \in [q] \\
&(b) P_{\phi(n_j)} \oplus P_{\phi(t_j)} \neq \lambda'_j (\neq \mathbf{0}),\ \forall j \in [q+1, q+r] \\
&(c) P_{\phi(n_j)} \neq \lambda''_j (\neq \mathbf{0}),\ \forall j \in [q+r+1, q+v]
\end{aligned}$$

over the set of α many unknown variables $\mathcal{P} = \{P_1, \ldots, P_\alpha\}$ such that P_a may be equals to some $P_{\phi(n_i)}$ or $P_{\phi(t_i)}$, where $a \in \{\phi(n_j), \phi(t_j)\}, j \in [q+1, q+v]$. Now, if

- *(i) (ϕ', Λ') is good and*
- *(ii) $\xi_{\max} = 3$*

then the number of solutions for $\mathcal{P}$, denoted by $h_{\frac{3q}{2}}$ such that $P_i \neq P_j$ for all distinct $i, j \in \{1, \ldots, \alpha\}$ is

$$h_{\frac{3q}{2}} \geq \frac{(2^n)_{\frac{3q}{2}}}{2^{nq}} \left(1 - \frac{5q^3}{2^{2n}} - \frac{v}{2^n}\right). \tag{5}$$

An Overview of the Proof. We prove the result by induction on the number of blocks. Let us assume that we have exactly u many blocks with total number of equations $2u$ and total number of variables $3u$. Let us also assume that we have ω_{3u} many uni-variate and bi-variate affine non equations and h_{3u} be the total number of solutions of such a system of affine equations and non-equations. Now, we introduce one more block that adds two new equations and three new variables with the existing set of equations and variables. We also consider an affine bi-variate non-equation such that one of its variables comes from the existing $3u$ many variables and the other one comes from the newly introduced three variables. Moreover, we also consider an affine uni-variate non-equation such that the variable comes from the newly introduced three variables. Given we have h_{3u} many solutions, we calculate the number of solutions for the newly introduced affine equations and non-equations such that the solution does not end up with any inconsistencies.

Proof. As mentioned, our proof is an inductive proof based on the number of blocks u. Our first observation is that as (ϕ', Λ') is good, ϕ is independent and thus $\xi_{\max} = \zeta_{\max} + 1$ and hence, the maximum number of variables P_i that can reside in the same block is 3. For the simplicity of the proof, assume that we have exactly 3 variables at each blocks. Now, it is easy to see that Eqn. (5) holds when $u = 1$.

As the next step of the proof, let h_{3u} be the solutions for first $2u$ many affine equations, which we denote as $\mathcal{E}^{=}_{2u}$. Now as soon as we add the $(u+1)^{th}$ block, we consider the following bi-variate affine equations

$$P_{3u+1} \oplus P_{3u+2} = \lambda_{2u+1}, P_{3u+1} \oplus P_{3u+3} = \lambda_{2u+2}$$

and those bi-variate affine non-equations which are of the form

$$P_{\sigma_i} \oplus P_{\delta_i} \neq \lambda'_i, \text{ where } \sigma_i \in \{1, \ldots, 3u+3\}, \delta_i \in \{3u+1, 3u+2, 3u+3\},$$

and also those uni-variate affine non-equations of the form

$$P_{\delta_i} \neq \lambda''_i, \text{ where } \delta_i \in \{3u+1, 3u+2, 3u+3\}.$$

Let v' and v'' be the number of such bi-variate and uni-variate affine non-equations. Now, note that each such bi-variate affine non-equation of the form $P_{\sigma_i} \oplus P_{\delta_i} \neq \lambda'_i$ where $\sigma_i \in \{1, \ldots, 3u+3\}, \delta_i \in \{3u+1, 3u+2, 3u+3\}$ can be written as $P_{3u+1} \neq P_{\sigma_i} \oplus \lambda^\star_i$, where $\sigma_i \in \{1, \ldots, 3u+3\}$ and $\lambda^\star_i \in \{\lambda'_i, \lambda'_i \oplus \lambda_{2u+1}, \lambda'_i \oplus \lambda_{2u+2}\}$. Moreover, each such uni-variate affine non-equation of the form $P_{\delta_i} \neq \lambda''_i$ where $\delta_i \in \{3u+1, 3u+2, 3u+3\}$ can be written as $P_{3u+1} \neq \lambda^{\star\star}_i$, where $\lambda^{\star\star}_i \in \{\lambda''_i, \lambda''_i \oplus \lambda_{2u+1}, \lambda''_i \oplus \lambda_{2u+2}\}$.

Now h_{3u+3} counts for the number of solutions to $\{P_1, \ldots, P_{3u}, P_{3u+1}, P_{3u+2}, P_{3u+3}\}$ such that

- $\{P_1, \ldots, P_{3u}\}$ is a valid solution of $\mathcal{E}^{=}_{2u}$.
- $P_{3u+1} \oplus P_{3u+2} = \lambda_{2u+1}, P_{3u+1} \oplus P_{3u+3} = \lambda_{2u+2}$.
- $P_{3u+1} \notin \{P_1, \ldots, P_{3u}, P_1 \oplus \lambda_{2u+1}, \ldots, P_{3u} \oplus \lambda_{2u+1}, P_1 \oplus \lambda_{2u+2}, \ldots, P_{3u} \oplus \lambda_{2u+2}\}$.
- $P_{3u+1} \notin \{P_{\sigma_1} \oplus \lambda^\star_1, \ldots, P_{\sigma_{v'}} \oplus \lambda^\star_{v'}\}$.
- $P_{3u+1} \notin \{\lambda^{\star\star}_1, \ldots, \lambda^{\star\star}_{v''}\}$.

Let $V_1 = \{P_1, \ldots, P_{3u}\}, V_2 = \{P_1 \oplus \lambda_{2u+1}, \ldots, P_{3u} \oplus \lambda_{2u+1}\}$, $V_3 = \{P_1 \oplus \lambda_{2u+2}, \ldots, P_{3u} \oplus \lambda_{2u+2}\}$, $V_4 = \{P_{\sigma_1} \oplus \lambda^\star_1, \ldots, P_{\sigma_{v'}} \oplus \lambda^\star_{v'}\}$ and $V_5 = \{\lambda^{\star\star}_1, \ldots, \lambda^{\star\star}_{v''}\}$. Note that, $|V_i| = 3u, i = 1, 2, 3$ and $|V_4| = v'$, $|V_5| = v''$. Therefore, we can write

$$\begin{aligned} h_{3u+3} &= h_{3u}(2^n - |V_1 \cup V_2 \cup V_3 \cup V_4 \cup V_5|) \geq h_{3u}(2^n - |V_1| - |V_2| - |V_3| - |V_4| - |V_5|) \\ &\geq h_{3u}(2^n - 9u - v' - v''). \end{aligned}$$

By applying repeated induction, we obtain

$$h_{\frac{3q}{2}} \geq \left(2^n - 9(\frac{q}{2} - 1) - v' - v''\right) h_{3(\frac{q}{2}-1)} \geq \ldots \geq \prod_{u=0}^{q/2-1} (2^n - 9u - v' - v'')$$

for which we have,

$$\begin{aligned}
\frac{h_{\frac{3q}{2}}2^{nq}}{(2^n)_{\frac{3q}{2}}} &\geq \prod_{u=0}^{q/2-1} \frac{2^{2n}(2^n-9u-v'-v'')}{(2^n-3u)(2^n-3u-1)(2^n-3u-2)} \\
&\geq \prod_{u=0}^{q/2-1} \frac{2^{2n}(2^n-9u-v'-v'')}{2^{3n}-(9u+3)2^{2n}+(27u^2+18u+2)2^n} \\
&\overset{[1]}{\geq} \prod_{u=0}^{q/2-1} \left(1+\frac{3}{2^n}-\frac{27u^2+18u+2}{2^{2n}}-\frac{v'+v''}{2^n}\right) \\
&\overset{[2]}{\geq} \prod_{u=0}^{q/2-1} \left(1-\frac{27u^2}{2^{2n}}-\frac{9u^2}{2^{2n}}-\frac{v'+v''}{2^n}\right) \geq \prod_{u=0}^{q/2-1} \left(1-\frac{36u^2}{2^{2n}}-\frac{v'+v''}{2^n}\right) \\
&\geq \left(1-\sum_{u=0}^{q/2-1}\frac{36u^2}{2^{2n}}-\sum_{u=0}^{q/2-1}\frac{v'+v''}{2^n}\right) \overset{[3]}{\geq} \left(1-\frac{5q^3}{2^{2n}}-\frac{v}{2^n}\right)
\end{aligned}$$

where [1] follows from the assumptions $u \leq 2^n/9$, [2] follows as $\frac{9u^2}{2^{2n}} \geq \frac{(18u+3)}{2^{2n}} - \frac{3}{2^n}$ and [3] follows as $\sum_{u=0}^{q/2-1}(v'+v'') \leq v$. □

4 DWCDM and Its Security Result

In this section, we discuss our proposed construction DWCDM and state its security in nonce respecting and nonce misuse setting. Let us recall the DWCDM construction:

$$\mathsf{DWCDM}[\mathsf{E},\mathsf{E}^{-1},\mathsf{H}](N,M) := \mathsf{E}_K^{-1}(\mathsf{E}_K(N) \oplus N \oplus \mathsf{H}_{K_h}(M))$$

where $N = N^*\|0^{n/3}$. E_K is a n-bit block cipher and H_{K_h} is an ϵ_1-regular, ϵ_2-AXU and ϵ_3-3-way regular n-bit keyed hash function. A schematic diagram of DWCDM is shown in Fig.4.1. Note that, DWCDM is structurally similar to EWCDM, but unlike EWCDM, our construction uses the same block cipher key and the last block cipher call of EWCDM is replaced by its decryption function. Moreover, DWCDM cannot exploit the full nonce space like EWCDM, otherwise its beyond birthday security will be compromised as explained below.

4.1 Why DWCDM Cannot Accommodate Full n-bit Nonce

As mentioned above, for DWCDM we need to reduce the nonce space to $2n/3$-bits. If it uses the full nonce space then using a nonce respecting adversary A who set the tags as nonce repeatedly, can mount a birthday bound forging attack on DWCDM as follows:

Suppose, an adversary starts with query (N, M) and then makes a chain of queries of the form (T_{i-1}, M) where (T_{i-1}, M) is the i-th query and T_{i-1} is the

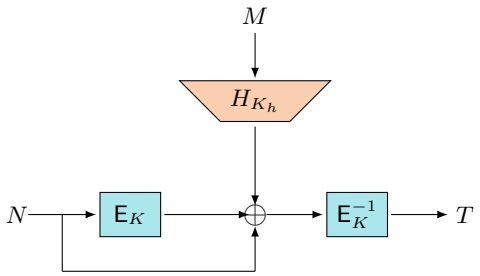

Fig. 4.1. Decrypted Wegman-Carter with Davies-Meyer Construction.

response of the previous $(i-1)$-th query, until the first time collision occurs (i.e. a responce matches with one of the previous responses). If the adversary makes upto $q \approx 2^{n/2}$ queries, it gets a collision $T_i = T_j$ with high probability. Interestingly, if $(j-i)$ [10] is even (which holds with probability $1/2$), then

$$T_j = T_i \text{ iff } (T_i + T_{i+1} + \cdots + T_{j-1} = 0).$$

Now, this property can be easily used A to predict $T_i = T_j$ if it finds $T_i + T_{i+1} + \cdots + T_{j-1} = 0$ for some i, j such that $(j-i)$ is even. Thus, A can mount the following forging attack as shown in Figure 4.2.

MAC Adversary A

1. Take any arbitrary N and M; set $T_0 \leftarrow N$.
2. **for** $(j = 1; j \leq q; j++)$
3. $\quad T_j \leftarrow \mathsf{DWCDM}(T_{j-1}, M)$.
4. $\quad$ **for**$(i = j-1; i \geq 0; i = i-2)$
5. $\quad\quad$ **if**$((T_i + \cdots + T_j) = 0)$
6. $\quad\quad\quad$ **forge**(T_j, M, T_i).

Fig. 4.2. Birthday bound MAC attack against DWCDM if full nonce space is used.

However, if we restrict the nonce space to $2n/3$ bits, then this attack doesn't work because now using the tag as a valid nonce is a probabilistic event. Probability that a tag is a valid nonce is $2^{-n/3}$. This restricts the adversary from forming a chain as used in the attack. In fact, if adversary makes $2^{2n/3}$ many MAC queries then the expected number of tags whose last $n/3$ bits are all zeros is $2^{n/3}$. Now, if adversary uses these $2^{n/3}$ tags as the nonces, then the expected number of tags whose last $n/3$ bits are zeros is 1 and then adversary cannot

[10] we assume $j > i$.

proceed further. This phenomenon effectively invalidates the above attack to happen.

4.2 Nonce Respecting Security of DWCDM

In this section, we state that DWCDM is secure up to $2^{2n/3}$ MAC queries and 2^n verification queries against nonce respecting adversaries. Formally, the following result bounds the MAC advantage of DWCDM against nonce respecting adversaries.

Theorem 2. *Let $\mathcal{M}, \mathcal{K}$ and $\mathcal{K}_h$ be finite and non-empty sets. Let $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $\mathsf{H} : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ϵ_2-AXU, ϵ_3-3-way regular and ϵ_1-regular hash function. Then, the MAC advantage for any (q_m, q_v, t) nonce respecting adversary against $\mathsf{DWCDM}[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$ is given by,*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{DWCDM}[\mathsf{E},\mathsf{E}^{-1},\mathsf{H}]}(q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m + q_v, t') + \frac{2q_m}{2^{2n/3}} + q_m\epsilon_1 + \frac{2q_m\epsilon_2}{2^{n/3}} + 2q_v.\max\{\epsilon_1, \epsilon_2, \epsilon_3\} + \frac{(q_m + q_v)}{2^n} + \frac{5q_m^3}{2^{2n}},$$

where $t' = O(t + (q_m + q_v)t_H)$, t_H be the time for computing the hash function. By assuming $\epsilon_1, \epsilon_2, \epsilon_3 \approx 2^{-n}$ and $q_m \leq 2^{2n/3}$, DWCDM is secured up to roughly $q_m \approx 2^{2n/3}$ MAC queries and $q_v \approx 2^n$ verification queries.

4.3 Nonce Misuse Security of DWCDM

Similar to EWCDM [15], one can prove that $\mathsf{DWCDM}[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$ is birthday bound secure MAC against nonce misuse adversaries. In particular, DWCDM is secure up to $2^{n/2}$ MAC queries and 2^n verification queries against nonce misuse adversaries and that the security bound is essentially tight. More formally, we have the following MAC security result of DWCDM in nonce misuse setting.

Theorem 3. *Let $\mathcal{M}, \mathcal{K}$ and $\mathcal{K}_h$ be finite and non-empty sets, $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $\mathsf{H} : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ϵ_1-regular and ϵ_2-AXU hash function. Then, the MAC security of $\mathsf{DWCDM}[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$ in nonce misuse setting is given by*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{DWCDM}}(q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m + q_v, t') + q_m^2\epsilon_2 + \frac{4q_m^2}{2^n} + q_m\epsilon_1 + \frac{(q_m + q_v)}{2^n},$$

where $t' = O(t + q(q_m + q_v)t_H), t_H$ be the time for computing hash function.

By assuming $\epsilon_1 \approx 2^{-n}$ and $\epsilon_2 \approx 2^{-n}$, DWCDM is secure up to roughly $q_m \approx 2^{n/2}$ MAC queries and $q_v \approx 2^n$ verification queries. For completeness of the paper, we present the proof of Theorem 3 in Appendix A.

Tightness of the Bound

We show that the above bound of DWCDM is tight by demonstrating a forging attack which shows thats roughly $2^{n/2}$ MAC queries are enough to break the MAC security of DWCDM when an adversary is allowed to repeat nonce only for once. The attack is as follows:

1. Adversary A makes q many MAC queries (N_i, M_i) with distinct nonces where a collision in the response, i.e. $T_i = T_j$ for some $i < j$ occurs.
2. Make a MAC query (N_j, M_i). Let T_{q+1} be the response.
3. Forge with (N_i, M_j, T_{q+1}).

As $\Pi(T_{q+1}) = \Pi(N_i) \oplus N_i \oplus \mathsf{H}_{K_h}(M_j)$, (N_i, M_j, T_{q+1}) is a valid forgery. If we make $q = 2^{n/2}$ many queries, with very high probability, we will get a collision in step 1, and mount the attack. Note that, the attack does not exploit any specific properties of the hash function and a single time repitition of nonce makes the construction vulnerable above birthday bound security.

4.4 nPolyMAC: An Instantiation of DWCDM

In this section, we propose nPolyMAC, an algebraic hash function based instantiation of DWCDM, as defined in Eqn. (3), as the underlying hash function of DWCDM construction.

PolyHash [26] is one of the popular examples of algebraic hash function. For a hash key K_h and a message $M = (M_1, M_2, \ldots, M_l)$ such that for all $i = 1, 2, \ldots l-1$ $|M_i| = n$ and $|M_l| \leq n$, we define the PolyHash as follows:

$$\mathsf{Poly}_{K_h}(M_1, M_2, \ldots, M_l) = (M_l \| 0^{n-|M_l|}) \cdot K_h \oplus M_{l-1} \cdot K_h^2 \oplus \ldots \oplus M_1 \cdot K_h^l.$$

It has already been shown in Proposition 1 that Poly is $\ell/2^n$ 3-way and 1-way regular hash function. Moreover, the AXU advantage of Poly is $\ell/2^n$. Following these results, we show in the following that $\mathsf{nPolyMAC}[\mathsf{Poly}, \mathsf{E}, \mathsf{E}^{-1}]$ is secure up to $2^{2n/3}$ MAC and 2^n verification queries against nonce respecting adversaries.

Theorem 4. *Let $\mathcal{K}, \mathcal{K}_h$ and $\mathcal{M}$ be three non-empty finite sets. Let $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher. Then, the MAC security of* nPolyMAC *in nonce respeting setting is given by*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{nPolyMAC}}(q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m + q_v, t') + \frac{11 q_m \ell}{2^{2n/3}} + \frac{3 q_v \ell}{2^n},$$

where $t' = O(t + (q_m + q_v)\ell)$, ℓ be the maximum number of message blocks among all q queries.

The proof of the theorem directly follows from Proposition 1 and Theorem 2 with the assumption $q_m \leq 2^{2n/3}$.

5 Proof of Theorem 2

In this section, we prove Theorem 2. We would like to note that we will often refer to the construction $\mathsf{DWCDM}[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$ as simply DWCDM where the underlying primitives are assumed to be understood.

As the first step of the proof, we replace E_K and E_K^{-1} with an n-bit uniform random permutation Π and its inverse Π^{-1} respectively both in the MAC and verification oracle of DWCDM and denote the construction as $\mathsf{DWCDM}^*[\Pi, \Pi^{-1}, \mathsf{H}]$. This conversion will add a term $\mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m + q_v, t')$ where $t' = O(t + (q_m + q_v)t_H)$ and hence one simply has

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{DWCDM}}(q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m + q_v, t') + \underbrace{\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{DWCDM}^*}(q_m, q_v, t)}_{\delta^*}. \quad (6)$$

Now, our goal is to upper bound δ^*. For doing this, we consider $\mathsf{TG}[\Pi, \Pi^{-1}, \mathsf{H}_{K_h}]$ be the tag generation and $\mathsf{VF}[\Pi, \Pi^{-1}, \mathsf{H}_{K_h}]$ be the verification oracle of DWCDM^* construction. We also consider that Rand be a perfect random oracle that on input $(N, M) \in \{0,1\}^n \times \mathcal{M}$, returns T, sampled uniformly at random from $\{0,1\}^n$, whereas Rej be an oracle with inputs (N, M, T), returns always $\perp$ (i.e. rejects). Now, due to [15, 20] we write

$$\delta^* := \max_{\mathsf{D}} \Pr[\mathsf{D}^{\mathsf{TG}[\Pi,\Pi^{-1},\mathsf{H}_{K_h}],\mathsf{VF}[\Pi,\Pi^{-1},\mathsf{H}_{K_h}]} = 1] - \Pr[\mathsf{D}^{\mathsf{Rand},\mathsf{Rej}} = 1],$$

where the maximum is taken over all non-trivial distinguishers D. This formulation allows us to apply the H-Coefficient Technique [30], as we explain in more detail below, to prove

$$\delta^* \leq \frac{2q_m}{2^{2n/3}} + q_m\epsilon_1 + \frac{2q_m\epsilon_2}{2^{n/3}} + 2q_v.\max\{\epsilon_1, \epsilon_2, \epsilon_3\} + \frac{(q_m + q_v)}{2^n} + \frac{5q_m^3}{2^{2n}}. \quad (7)$$

H-Coefficient Technique. From now on, we fix a non-trivial distinguisher D that interacts with either (1) the real oracle $(\mathsf{TG}[\Pi, \Pi^{-1}, \mathsf{H}_{K_h}], \mathsf{VF}[\Pi, \Pi^{-1}, \mathsf{H}_{K_h}])$ for a random permutation Π, its inverse Π^{-1} and a random hashing key K_h or (2) the ideal oracle $(\mathsf{Rand}, \mathsf{Rej})$ making at most q_m queries to its left (MAC) oracle and at most q_v queries to its right (verification) oracle, and outputting a single bit. We let

$$\mathbf{Adv}(\mathsf{D}) = \Pr[\mathsf{D}^{\mathsf{TG}[\Pi,\Pi^{-1},\mathsf{H}_{K_h}],\mathsf{VF}[\Pi,\Pi^{-1},\mathsf{H}_{K_h}]} = 1] - \Pr[\mathsf{D}^{\mathsf{Rand},\mathsf{Rej}} = 1].$$

We assume that D is computationally unbounded and hence wlog deterministic and that it never repeats a query. Let

$$\tau_m := \{(N_1, M_1, T_1), (N_2, M_2, T_2), \ldots, (N_{q_m}, M_{q_m}, T_{q_m})\}$$

be the list of MAC queries of D and its corresponding responses. Note that, as D is nonce respecting, there cannot be any repetition of triplet in τ_m. Let also

$$\tau_v := \{(N_1', M_1', T_1', b_1'), (N_2', M_2', T_2', b_2'), \ldots, (N_{q_v}', M_{q_v}', T_{q_v}', b_{q_v}')\}$$

be the list of verification queries of D and its corresponding responses, where for all j, $b'_j \in \{\top, \bot\}$ denotes the accept ($b'_j = \top$) or reject ($b'_j = \bot$). The pair (τ_m, τ_v) constitutes the query transcript of the attack. For convenience, we slightly modify the experiment where we reveal to the distinguisher (after it made all its queries and obtains corresponding responses but before it output its decision) the hashing key K_h, if we are in the real world, or a uniformly random dummy key K_h if we are in the ideal world. All in all, the transcript of the attack is $\tau = (\tau_m, \tau_v, K_h)$ where τ_m and τ_v is the tuple of MAC and verification queries respectively. We will often simply name a tuple $(N, M, T) \in \tau_m$ a *MAC query*, and a tuple $(N', M', T', b) \in \tau_v$ a *verification query*.

A transcript τ is said to be an *attainable* (with respect to D) transcript if the probability to realize this transcript in ideal world is non-zero. For an attainable transcript $\tau = (\tau_m, \tau_v, K_h)$, any verification query $(N'_i, M'_i, T'_i, b'_i) \in \tau_v$ is such that $b'_i = \bot$. We denote Θ to be the set of all attainable transcripts and X_{re} and X_{id} denotes the probability distribution of transcript τ induced by the real world and ideal world respectively. In the following we state the main lemma of the H-coefficient technique (see e.g. [13] for the proof).

Lemma 1. *Let D be a fixed deterministic distinguisher and $\Theta = \Theta_{\mathrm{g}} \sqcup \Theta_{\mathrm{b}}$ (disjoint union) be some partition of the set of all attainable transcripts. Suppose there exists $\epsilon_{\mathrm{ratio}} \geq 0$ such that for any $\tau \in \Theta_{\mathrm{g}}$,*

$$\frac{\Pr[X_{\mathrm{re}} = \tau]}{\Pr[X_{\mathrm{id}} = \tau]} \geq 1 - \epsilon_{\mathrm{ratio}},$$

and there exists $\epsilon_{\mathrm{bad}} \geq 0$ such that $\Pr[X_{\mathrm{id}} \in \Theta_{\mathrm{b}}] \leq \epsilon_{\mathrm{bad}}$. Then, $\mathbf{Adv}(\mathsf{D}) \leq \epsilon_{\mathrm{ratio}} + \epsilon_{\mathrm{bad}}$.

The remaining of the proof of Theorem 2 is structured as follows: in section 5.1 we define the transcript graph; in section. 5.2 we define bad transcripts and upper bound their probability in the ideal world; in section 5.3, we analyze good transcripts and prove that they are almost as likely in the real and the ideal world. Theorem 2 follows easily by combining Lemma 1, Eqn. (6) and (7) above, and Lemmas 3 and 4 proven below.

5.1 Transcript Graph

Given a transcript $\tau = (\tau_m, \tau_v, K_h)$, we define the following two types of graphs: (a) MAC Graph and (b) Verification Graph.

MAC Graph. Given a transcript $\tau = (\tau_m, \tau_v, K_h)$, we define the *MAC graph*, denoted as $G^{\mathtt{m}}_{\tau}$ as follows:

$$G^{\mathtt{m}}_{\tau} = ([q_m], E^{\mathtt{m}}) \text{ where } E^{\mathtt{m}} = \{(i,j) \in [q_m] \times [q_m] : N_i = T_j \vee N_j = T_i \vee T_i = T_j\}.$$

For the sake of convenience, we denote the edge (i, j) as a dotted line when $T_i = T_j$, else we denote it as a continuous line. Thus, the edge set of $G^{\mathtt{m}}_{\tau}$ consists

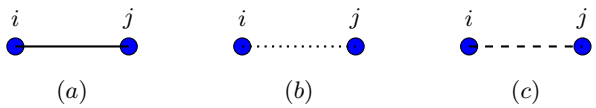

Fig. 5.1. Different types of edges of MAC and Verification Graphs. $(a) : N_i = T_j/T_i = N_j$, $(b) : T_i = T_j$, $(c) : N_i = N_j$.

of two different types of edges as depicted in Fig. 5.1 (a) and (b). Note that, for a MAC graph we cannot have edges of type (c).
Given such a MAC graph, we can partition the set of vertices in the following way: if vertex i and j are connected by an edge then they belong to the same partition. Each partition is called a component of the graph and the number of vertices in the component is called its size, which we denote as ζ.

Verification Graph. Given a MAC graph $G^{\mathtt{m}}_{\tau}$, we define *Verification graph*, denoted as $G^{\mathtt{v}}_{\tau}$, by extending $G^{\mathtt{m}}_{\tau}$ with adding one more vertex and at most two edges for incorporating a verification query as follows: For convenience, we reorder the set of MAC queries and verification queries so that all verification queries appears after all MAC queries. Therefore, after such a reordering, j-th verification query becomes $(q_m + j)$-th verification query. Let $(q_m + j)$-th verification query be $(N'_{q_m+j}, M'_{q_m+j}, T'_{q_m+j}, b'_{q_m+j}) \in \tau_v$ and $G^{\mathtt{m}}_{\tau}$ be the MAC graph corresponding to $\tau = (\tau_m, \tau_v, K_h)$. Then we define $G^{\mathtt{v}}_{\tau} = ([q_m] \cup \{q_m + j\}, E^{\mathtt{v}})$ where $E^{\mathtt{v}}$ is defined as follows:

$E^{\mathtt{v}} = E^{\mathtt{m}} \cup \{(q_m+j, r), (q_m+j, s) : r \neq s \in [q_m]$ such that either of (1)-(4) holds$\}$.

$$\begin{cases} (1)\ N'_{q_m+j} = N_r \wedge T'_{q_m+j} = N_s \\ (2)\ N'_{q_m+j} = N_r \wedge T'_{q_m+j} = T_s \\ (3)\ N'_{q_m+j} = T_r \wedge T'_{q_m+j} = N_s \\ (4)\ N'_{q_m+j} = T_r \wedge T'_{q_m+j} = T_s \end{cases}$$

Definition 3 (Valid Cycle). *A cycle $C = (i_1, i_2, \ldots, i_p)$ of length p in the MAC graph $G^{\mathtt{m}}_{\tau}$ is said to be valid if the imposed equality pattern of (N, T), generated out of C, derives*

$$\mathbf{0} = \bigoplus_{i \in C} \left(N_i \oplus \mathsf{H}_{K_h}(M_i) \right)$$

equation from the given system of equations.

Similar to the definition of valid cycle of MAC graph, one can define the valid cycle for the Verification graph also. Note that, the definition of valid cycle in MAC graph or verification graph actually resembles to the alternating cycle as stated in section. 3.1. Now, we make an important observations about the MAC queries (in ideal oracle) as follows:

Lemma 2. *For two MAC queries i, j, we have*

$$(a)\ \textit{if } i < j,\ \Pr[T_j = N_i] = \frac{1}{2^n}; \quad (b)\ \textit{if } i > j,\ \Pr[T_j = N_i] = \frac{1}{2^{n/3}}.$$

Proof. Proof of the first result holds due to the randomness of T_j, i.e a randomly sampled value T_j is equal to a fixed nonce value N_i holds with probability 2^{-n}. For the later one, condition $i > j$ ensures that one can set the nonce value N_i to a previously sampled tag value T_j. But this would be valid only when the last $n/3$ bits of T_i are all zero, probability of which is $2^{-n/3}$. □

5.2 Definition and Probability of Bad Transcripts

In this section, we define and bound the probability of bad transcript in ideal world. But, before that we first briefly justify the reason about our identified bad events and there after we define the bad transcript accordingly.

Let $\tau = (\tau_m, \tau_v, K_h)$ be an attainable transcript. Then, for all MAC queries (N_i, M_i, T_i) in real oracle, we have

$$i \in \{1, \ldots, q_m\}, \mathsf{\Pi}(N_i) \oplus \mathsf{\Pi}(T_i) = N_i \oplus \mathsf{H}_{K_h}(M_i).$$

Moreover, for all verification queries (N'_a, M'_a, T'_a, b_a) in real oracle, we have

$$a \in \{1, \ldots, q_v\}, \mathsf{\Pi}(N'_a) \oplus \mathsf{\Pi}(T'_a) \neq N'_a \oplus \mathsf{H}_{K_h}(M'_a).$$

We refer to the system of equations as "MAC Equations" which involve only the MAC queries. Similarly, we refer to the system of non-equations as "Verification non-equations" which involve only the verification queries.

Therefore, from a given attainable transcript τ, one can write exactly q_m many affine equations and q_v many non-equations. Now, as one needs to lower bound the number of solutions of this system of equations and non-equations (for analyzing the real interpolation probability), it essentially leads us to the model of extended Mirror theory where the equivalence of two set up is established as follows:

$$\begin{cases} \phi'(n_i) = N_i,\ \phi'(t_i) = T_i,\ \lambda_i = N_i \oplus \mathsf{H}_{K_h}(M_i),\ i \in \{1, \ldots, q_m\} \\ \phi'(n_a) = N'_a,\ \phi'(t_a) = T'_a,\ \lambda'_a = N'_a \oplus \mathsf{H}_{K_h}(M'_a),\ a \in \{1, \ldots, q_v\} \end{cases}$$

Recall that, (ϕ', Λ') where $\Lambda' = (\lambda_1, \ldots, \lambda_{q_m}, \lambda'_1, \ldots, \lambda'_{q_v})$, was characterized to be bad if either of the following holds:

(i) $\phi(n_i) = \phi(t_i)$.

(ii) - $\phi(n_i) = \phi(n_j)$ and $\phi(t_i) = \phi(t_j)$
- $\phi(n_i) = \phi(t_j)$ and $\phi(t_i) = \phi(n_j)$ for $i \neq j \in [q_m]$.

(iii) there is an alternating cycle.

(*iv*) for all $j \in [q_v]$ and $i_1, \ldots, i_c \in [q_m]$, $c \geq 0$, such that $\{i_1, \ldots, i_c, q_m + j\}$ is dependent system then $\lambda_{i_1} \oplus \cdots \oplus \lambda_{i_c} \oplus \lambda'_j = \mathbf{0}$.

where $\phi = \phi'_{|q_m}$. Therefore, with the help of equivalence of two set up as established above, we justify our identified bad events:

- $(i) \Rightarrow N_i = T_i$
- $(ii) \Rightarrow$ existence of a valid cycle in the MAC graph $G^{\mathtt{m}}_{\tau}$.
- $(iii) \Rightarrow N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j), T_i = T_j$ or $N_i = T_j, N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j)$ such that $i \neq j \in [q_m]$.

Moreover, recall that while considering the non-equation then we considered that any of q_v non-equations can be determined from a subset of q_m many affine equations with their correponding sum of λ constant becomes zero, which is to say that

- the verification graph $G^{\mathtt{v}}_{\tau}$ contains any valid cycle.

Summarizing above, we now define the bad transcript.

Definition 4. *A transcript $\tau = (\tau_m, \tau_v, K_h)$ is said to be* bad *if the associated MAC graph $G^{\mathtt{m}}_{\tau}$ and the Verification graph $G^{\mathtt{v}}_{\tau}$ satisfies the either of the following properties:*

- $\mathsf{B0}$: *$\exists i \in [q_m]$ such that $T_i = \mathbf{0}$.*
- $\mathsf{B1}$: *$G^{\mathtt{m}}_{\tau}$ has a component of size 3 or more.*
- $\mathsf{B2}$: *$G^{\mathtt{m}}_{\tau}$ contains a valid cycle of any arbitrary length that also includes the self loop (that implicitly takes care of the condition $N_i = T_i$).*
- $\mathsf{B3}$: *$G^{\mathtt{v}}_{\tau}$ contains a valid cycle of any arbitrary length that involves the verification query.*

Moreover, τ is also said to be bad if

- $\mathsf{B4}$: *$\exists i \neq j \in [q_m]$ such that $N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j), T_i = T_j$.*
- $\mathsf{B5}$: *$\exists i \neq j \in [q_m]$ such that $N_i = T_j, N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j)$.*
- $\mathsf{B6}$: *$\exists i \in [q_m]$ such that $\mathsf{H}_{K_h}(M_i) = N_i$.*

Condition $\mathsf{B1}$ actually imposes a restriction on the block maximality as we do not allow to have a larger component size for a good transcript. Condition $\mathsf{B6}$ ensures that for a good transcript, all the elements of the tuple $\big(N_1 \oplus \mathsf{H}_{K_h}(M_1), \ldots, N_{q_m} \oplus \mathsf{H}_{K_h}(M_{q_m})\big)$ are non-zero. Note that, if we do not consider the condition $\mathsf{B6}$, then for a good attainable transcript the real interpolation probability would become zero.

We denote $\Theta_b \subseteq \Theta$ be the set of all attainable bad transcripts and the event B denotes $\mathsf{B} := \mathsf{B0} \vee \mathsf{B1} \vee \mathsf{B2} \vee \mathsf{B3} \vee \mathsf{B4} \vee \mathsf{B5} \vee \mathsf{B6}$, We bound the probability of event B in ideal world as follows:

Lemma 3. *Let X_{id} and Θ_b be defined as above. If $q_m \leq 2^{2n/3}$ and $q_v \leq 2^n$, then*

$$\Pr[X_{\mathrm{id}} \in \Theta_b] \leq \epsilon_{\mathsf{bad}} = \frac{2q_m}{2^{2n/3}} + \frac{q_m}{2^n} + q_m\epsilon_1 + \frac{2q_m\epsilon_2}{2^{n/3}} + 2q_v.\max\{\epsilon_1, \epsilon_2, \epsilon_3\}.$$

Proof of this lemma is deferred to section 5.4.

5.3 Analysis of Good Transcripts

In this section, we show that for a good transcript τ, realizing τ is almost as likely in the real world as in the ideal world. Formally, we prove the following lemma.

Lemma 4. *Let $\tau = (\tau_m, \tau_v, K_h)$ be a good transcript. Then*

$$\frac{\mathsf{p}_{\mathrm{re}}(\tau)}{\mathsf{p}_{\mathrm{id}}(\tau)} := \frac{\Pr[X_{\mathrm{re}} = \tau]}{\Pr[X_{\mathrm{id}} = \tau]} \geq (1 - \epsilon_{\mathrm{ratio}}) = \left(1 - \frac{5q_m^3}{2^{2n}} - \frac{q_v}{2^n}\right).$$

Proof. Consider the good transcript $\tau = (\tau_m, \tau_v, K_h)$. Since in the ideal world the MAC oracle is perfectly random and the verification always rejects, one simply has

$$\mathsf{p}_{\mathrm{id}} := \Pr[X_{\mathrm{id}} = \tau] = \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{2^{nq_m}} \tag{8}$$

We must now lower bound the probability of getting τ in real world. We say that a permutation Π is compatible with τ_m if

$$\forall i \in [q_m], \Pi(N_i) \oplus \Pi(T_i) = \underbrace{N_i \oplus \mathsf{H}_{K_h}(M_i)}_{\lambda_i}$$

and we say that it is compatible with τ_v if

$$\forall a \in [q_v], \Pi(N'_a) \oplus \Pi(T'_a) \neq \underbrace{N'_a \oplus \mathsf{H}_{K_h}(M'_a)}_{\lambda'_a}.$$

We simply say that Π is compatible with τ if it is compatible with τ_m and τ_v. We denote $\mathsf{Comp}(\tau)$ the set of permutations that are compatible with τ. Therefore,

$$\begin{aligned}\mathsf{p}_{\mathrm{re}}(\tau) &= \frac{1}{|\mathcal{K}_h|} \cdot \Pr[\Pi \leftarrow\!\!\$\, \mathsf{Perm} : \Pi \in \mathsf{Comp}(\tau)] \\ &= \frac{1}{|\mathcal{K}_h|} \cdot \underbrace{\Pr[\Pi(N_i) \oplus \Pi(T_i) = \lambda_i, \forall i \in [q_m], \Pi(N'_a) \oplus \Pi(T'_a) \neq \lambda'_a, \forall a \in [q_v]]}_{\mathsf{P}_{mv}}\end{aligned}$$

Lower Bounding P_{mv}: Observe that lower bounding P_{mv} implies lower bounding the probability of the number of solutions to the following system of q_m many equations and q_v many non-equations:

$$(\mathcal{E}_m) = \begin{cases} \Pi(N_1) \oplus \Pi(T_1) = \lambda_1 \\ \Pi(N_2) \oplus \Pi(T_2) = \lambda_2 \\ \vdots \\ \Pi(N_{q_m}) \oplus \Pi(T_{q_m}) = \lambda_{q_m} \end{cases} \qquad (\mathcal{E}_v) = \begin{cases} \Pi(N'_1) \oplus \Pi(T'_1) \neq \lambda'_1 \\ \Pi(N'_2) \oplus \Pi(T'_2) \neq \lambda'_2 \\ \vdots \\ \Pi(N'_{q_v}) \oplus \Pi(T'_{q_v}) = \lambda'_{q_v} \end{cases}$$

Let us assume the distinct number of random variables in the above set of equations is α. As the transcript τ is good, we have the following properties:

- (i) all λ_i values are non-zero (otherwise condition $\mathsf{B6}$ is satisfied).
- (ii) (ϕ', Λ') is good.
- (iii) Finally, block maximality $\xi_{\max}$ is 3.

Above properties enable us directly to apply Theorem 1 to lower bound P_{mv} as follows:

$$\mathsf{P}_{mv} \geq \frac{1}{2^{nq_m}}\left(1 - \frac{5q_m^3}{2^{2n}} - \frac{q_v}{2^n}\right) \tag{9}$$

Therefore, from Eqn. (9), we have

$$\mathsf{p}_{\mathrm{re}}(\tau) \geq \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{2^{nq_m}} \cdot \left(1 - \frac{5q_m^3}{2^{2n}} - \frac{q_v}{2^n}\right) \tag{10}$$

Finally, taking the ratio of Eqn. (10) to Eqn. (8), the result follows. □

5.4 Proof of Lemma 3

In order to bound $\Pr[X_{\mathrm{id}} \in \Theta_b]$, it is enough to bound $\Pr[\mathsf{B}]$. Therefore, we write

$$\Pr[\mathsf{B}] \leq \sum_{\mathsf{v} \in \{0,1,4,5,6\}} \Pr[\mathsf{Bv}] + \Pr[\mathsf{B2} \mid \overline{\mathsf{B1}}] + \Pr[\mathsf{B3} \mid \overline{\mathsf{B0}} \wedge \overline{\mathsf{B1}} \wedge \overline{\mathsf{B2}}] \tag{11}$$

In the following, we bound the probabilities of all the bad events individually.

Bounding B0. Event $\mathsf{B0}$ occurs if there exists a MAC query whose response is all zero. For a fixed MAC query, the probability of this event in ideal world is 2^{-n} as the responses are sampled uniformly and independently to all other sampled random variables. Now, varying over all such MAC queries we obtain the bound to be

$$\Pr[\mathsf{B0}] \leq \frac{q_m}{2^n} \tag{12}$$

Bounding B1. Event $\mathsf{B1}$ occurs if there exists a component of size at least 3 in $G_\tau^{\mathtt{m}}$. This essentially implies there is a chain of two edges. Depending on whether the edges are dotted (Dot) or continuous (Con), there are three possible choices of components: components with both the edges being dotted, components with one dotted and one continuous edge and component with both continuous edges. We analyze each of them one by one:

(Dot-Dot) Component. $\exists i, j, k \in [q_m]$ such that $T_i = T_j = T_k$. For a fixed set of $i, j, k \in [q_m]$, this event is bounded by 2^{-2n} as each T_i is sampled uniformly at random from $\{0,1\}^n$. Summing over all possible choices of i, j and k, we obtain $\frac{q_m^3}{6 \cdot 2^{2n}}$ bound.

(Dot-Con) Component. $\exists i, j, k \in [q_m]$ such that $N_i = T_j = T_k$. For a fixed set of i, j and k, $T_j = T_k$ is bounded by 2^{-n}. Now, we have the following two sub cases: (a) if $j > i$ then from Lemma 2, probability of $N_i = T_j$ is bounded by

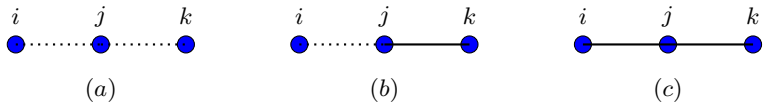

Fig. 5.2. Different components of size of three. $(a)T_i = T_j = T_k$, $(b)N_i = T_j = T_k$ or $T_i = N_j, T_j = T_k$ and $(c)N_i = T_j, N_j = T_k$ or $T_i = N_j, T_j = N_k$.

2^{-n} and thus the overall probability becomes 2^{-2n}. Summing over all possible choices of i, j, k, we obtain $\frac{q_m^3}{6 \cdot 2^{2n}}$. (b) if $j < i$ then from Lemma 2, probability of $N_i = T_j$ is bounded by $2^{-n/3}$. Hence the overall probability becomes $2^{-4n/3}$. In this case, choice of i is 1 and that of j and k is at most q_m. Therefore, summing over all possible choices of i, j, k we have $\frac{q_m^2}{2 \cdot 2^{4n/3}} \leq \frac{q_m}{2^{2n/3}}$, assuming $q_m \leq 2^{2n/3}$.
(Con-Con) Component. $\exists i, j, k \in [q_m]$ such that $N_i = T_j, N_j = T_k$. We bound this event using different sub cases. (a) when $i < j < k$, then due to Lemma 2, we obtain 2^{-2n} bound and varying over all possible choices of i, j, k, we obtain $\frac{q_m^3}{2^{2n}}$ bound. (b) when $i < j$ and $j > k$, then we obtain $2^{-4n/3}$ bound, but there is exactly one choice of j and q_m many choices for i and k. Hence, by summing over all possible choices of i, j, k we obtain $\frac{q_m^2}{2^{4n/3}} \leq \frac{q_m}{2^{2n/3}}$ bound. (c) $i > j, j < k$ is similar to case (b) and finally (d) when $i > j > k$, then due to Lemma 2, probability of the event is bounded by $2^{2n/3}$ and there is exactly one choice of i and j, leaving q_m choices for k which eventually gives $\frac{q_m}{2^{2n/3}}$ bound.

For each of the above cases, we obtain the maximum bound to be $\frac{q_m}{2^{2n/3}}$ and therefore, we have

$$\Pr[\mathsf{B1}] \leq \frac{q_m}{2^{2n/3}}. \tag{13}$$

Bounding B2 $|\overline{\mathsf{B1}}$: Recall that event B2 holds if there exists any cycle in $G_\tau^{\mathtt{m}}$. But, as we conditioned on $\overline{\mathsf{B1}}$, it is enough to bound the existence of a cycle of length one (self loop) and two (parallel edges).

Self Loop. A self loop or cycle of length 1 in $G_\tau^{\mathtt{m}}$ implies that $\exists i \in [q_m]$ such that $N_i = T_i$. For a fixed choice of i, probability of $N_i = T_i$ is bounded by 2^{-n} due to randomness of T_i. Summing over all choices of i, we obtain $\frac{q_m}{2^n}$ bound.

Fig. 5.3. (a) Self Loop: when $N_i = T_i$, (b) Parallel Edges: $N_i = T_j, N_j = T_i$.

Parallel Edges. A parallel edge or cycle of length 2 in $G_\tau^{\mathtt{m}}$ implies that $\exists i \neq j \in [q_m]$ such that $N_i = T_j, N_j = T_i$. For a fixed choice of i, j (without loss

of generality we assume $i < j$), probability of $N_i = T_j, N_j = T_i$ is bounded by $2^{-4n/3}$. This is because of Lemma 2, the probability of $N_i = T_j$ is bounded by 2^{-n} and the probability of $N_j = T_i$ is bounded by $2^{-n/3}$. As there exists only one choice of j and q_m many choices of i, summing over all possible choices of i and j, we obtain $\frac{q_m}{2^{4n/3}} \leq \frac{q_m}{2^{2n/3}}$ bound.

From the above two cases, we obtain the maximum bound to be $\frac{q_m}{2^{2n/3}}$ and thus we write

$$\Pr[\mathsf{B2} \mid \overline{\mathsf{B1}}] \leq \frac{q_m}{2^{2n/3}}. \tag{14}$$

Bounding B3 | $\overline{\mathsf{B0}} \wedge \overline{\mathsf{B1}} \wedge \overline{\mathsf{B2}}$. Recall that event B3 holds if there exists any cycle in G^{v}_{τ} and the sum of the corresponding $N \oplus \mathsf{H}_{K_h}(M)$ is zero. But, as we conditioned on $\overline{\mathsf{B0}} \wedge \overline{\mathsf{B1}} \wedge \overline{\mathsf{B2}}$, it is enough to bound the existence of a cycle of length one two and three.

Self Loop. A self loop or cycle of length 1 in G^{v}_{τ} implies that $\exists a \in [q_v]$ such that $N'_a = T'_a$ and $\mathsf{H}_{K_h}(M'_a) = N'_a$. Note that, for a fixed choice of a, the above event holds with probability at most ϵ_1 as we assume the hash function is ϵ_1 regular. Summing over all choices of a, we obtain $q_v\epsilon_1$ bound.

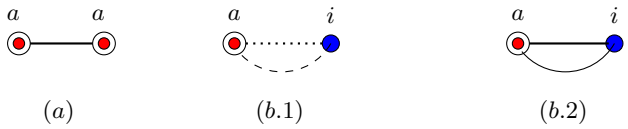

Fig. 5.4. (a) Self Loop: when $N'_a = T_a$, (b) Parallel Edges: ($b.1$) $N'_a = N_i, T'_a = T_i$, ($b.2$) $N'_a = T_i, T'_a = N_i$. Node with concentric circle denotes the verification query node.

Parallel Edges. A parallel edge or cycle of length 2 in G^{v}_{τ} implies that the edges would be (i) one dashed (Dash) and one dotted (Dot) or (ii) both continuous (Con-Con). Formally $\exists i \in [q_m], a \in [q_v]$:

$$\begin{cases} (\mathsf{Dot\text{-}Dash}) : N_i = N'_a, T_i = T'_a, \mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M'_a) = N_i \oplus N'_a \\ (\mathsf{Con\text{-}Con}) : N'_a = T_i, T'_a = N_i, \mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M'_a) = N_i \oplus N'_a \end{cases}$$

- Bounding (Dot-Dash) Edges. For a fixed choice of i and a, probability of $N_i = N'_a, T_i = T'_a, \mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M'_a) = N_i \oplus N'_a$ is bounded by ϵ_2 due to the randomness of hash key K_h (note that $M_i \neq M'_a$, as we have assumed a non-trivial distinguisher). As there exists only one choice of i for which the above probability is bounded by ϵ_2, summing over all possible choices of i and a, we obtain the bound $q_v\epsilon_2$.
- Bounding (Con-Con) Edges. Similarly, for a fixed choice of i, a, probability of $N'_a = T_i, T'_a = N_i, \mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M'_a) = N_i \oplus N'_a$ is bounded by ϵ_2 due to the randomness of hash key K_h, Note that, in this case there exists only

two choices of i (as we have at most two collision of T) for which the above probability is bounded by ϵ_2. Summing over all possible choices of i and a, we obtain the bound $2q_v\epsilon_2$.

Thus, we see from the above two cases that the probability of forming a parallel edge is bounded by $2q_v\epsilon_2$.

Closed Triangle. For the case of a closed triangle, it has the following restrictions on the edges: (a) maximum one edge can be dashed and if exist, it must contain the verification node a, (b) maximum one edge can be dotted (otherwise there exists a self loop or parallel edges). These restrictions impose the following properties on the closed triangle:

- if (i, j)-th edge is dotted, then the other edges must be one dashed and one continuous.
- if (i, j)-th edge is continuous, then there must be a dashed edge.

Hence, a closed triangle or cycle of length 3 in G^{v}_{τ} implies that $\exists i, j \in [q_m], a \in [q_v]$ such that either of the following holds:

$$\begin{cases} (\mathsf{Con\text{-}Dash\text{-}Dot}) : N_i = T_j, N'_a = N_j, T'_a = T_i \text{ and} \\ \qquad\qquad \mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) \oplus \mathsf{H}_{K_h}(M'_a) = T_j \\ (\mathsf{Con\text{-}Con\text{-}Con}) : N_i = T_j, T'_a = N_j, N'_a = T_i \text{ and} \\ \qquad\qquad \mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) \oplus \mathsf{H}_{K_h}(M'_a) = T_i \oplus T_j \oplus T'_a. \\ (\mathsf{Dot\text{-}Dash\text{-}Con}) : T_i = T_j, N'_a = N_j, T'_a = N_i \text{ and} \\ \qquad\qquad \mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) \oplus \mathsf{H}_{K_h}(M'_a) = T_i \end{cases}$$

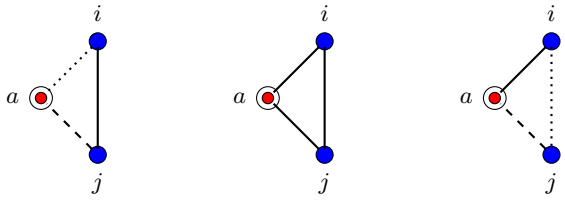

Fig. 5.5. Cycles of length 3 including the verification query which is denoted by the concentric circle node.

We bound each of these events for a fixed choice of i, j and a.

- Bounding (Con-Dash-Dot) Triangle. For a fixed choice of i, j, a, probability of the event is bounded by ϵ_3 using the randomness of the hash key as we have assumed the hash function is ϵ_3-3-way regular (note that we have conditioned on $\overline{\mathsf{B0}}$ and therefore $T_j \neq \mathbf{0}$). Note that, choice of j and i is one and two respectively. Hence, summing over all possible choices of indices, we obtain the bound to be $2q_v\epsilon_3$.

- Bounding (Con-Con-Con) Triangle. We analyze this case in two different subcases:
  - (a) For a fixed choice of i, j and a, if $N_j \neq T_i \oplus T_j$, then $\mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) \oplus \mathsf{H}_{K_h}(M'_a) \neq \mathbf{0}$ and thus we can bound the event by ϵ_3. In this case, choices of i and j is 2 an 1 respectively and hence we obtain the bound to be $2q_v\epsilon_3$.
  - (b) For a fixed choice of i, j and a, if $N_j = T_i \oplus T_j$, then $\mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) \oplus \mathsf{H}_{K_h}(M'_a) = \mathbf{0}$ and in that case we again consider two different sub cases: (i) if $i < j$, then $T_j = N_i$ holds with probability 2^{-n} and T_i is to be valid (i.e. last $n/3$ bits of T_i has to be zero), which holds with probability $2^{-n/3}$. Moreover, number of choices of i and j in this case is q_m and thus we obtain the bound to be $\frac{q_m^2}{2^{4n/3}} \leq \frac{q_m}{2^{2n/3}}$, assuming $q_m \leq 2^{2n/3}$. (ii) If $i > j$, then $T_j = N_i$ holds with probability $2^{-n/3}$ and T_i should be $N_i \oplus N_j$ which holds with probability 2^{-n}. In this case, number of choices of j is q_m and i in 1, resulting in probability of the event to be bounded by $\frac{q_m}{2^{4n/3}} \leq \frac{q_m}{2^{2n/3}}$.
- Bounding (Dot-Dash-Con) Triangle. For a fixed choice of i, j and a, the probability of the event is bounded by ϵ_3 as $\mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) \oplus \mathsf{H}_{K_h}(M'_a) = T_i$ holds with probability at most ϵ_3 (by the assumption that the hash function is ϵ_3-3-way regular). Note that, in this case choice of j and i is one. Hence, summing over all possible choices of indices, we obtain the bound to be $q_v\epsilon_3$.

Therefore, we see that for all the above cases, the maximum probability of forming a closed triangle in G^{v}_τ is $2q_v\epsilon_3$.

Therefore, we see from all of the above cases the maximum probability of forming a cycle in G^{v}_τ is $\max\{2q_v\epsilon_3, 2q_v\epsilon_2, q_v\epsilon_1\}$. Thus, we have

$$\Pr[\mathsf{B3} \mid \overline{\mathsf{B0}} \wedge \overline{\mathsf{B1}} \wedge \overline{\mathsf{B2}}] \leq \max\{2q_v\epsilon_3, 2q_v\epsilon_2, q_v\epsilon_1\} \tag{15}$$

Bounding B4: Recall that, the event $\mathsf{B4}$ holds if $\exists i \neq j \in [q_m]$ such that $N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j), T_i = T_j$. Since, in the ideal oracle the hash key is sampled independent to all previously sampled MAC responses T_i, we write

$$\Pr[\mathsf{B4}] \leq \sum_{i,j} \Pr[\mathsf{H}_{K_h}(M_i) \oplus \mathsf{H}_{K_h}(M_j) = N_i \oplus N_j] \cdot \Pr[T_i = T_j] \leq \frac{q_m^2 \cdot \epsilon_2}{2^n}. \tag{16}$$

Bounding B5: Recall that, the event $\mathsf{B5}$ holds if $\exists i \neq j \in [q_m]$ such that $N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j), N_i = T_j$. Now, we consider two sub cases:

- (a) for fixed i and j if $i < j$, then $N_i = T_j$ holds with probability 2^{-n} (due to Lemma 2) and $N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j)$ holds with probability ϵ_2. Summing over all possible choices of i and j we obtain the bound to be $\frac{q_m^2\epsilon_2}{2^n}$.
- (b) When $i > j$, then $N_i = T_j$ holds with probability $2^{-n/3}$ (due to Lemma 2) and as before $N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j)$ holds with probability ϵ_2. In

this case, possible choices of i and j is 1 and q_m respectively and therefore by summing over all possible choices of indices, we obtain the bound to be $\frac{q_m \epsilon_2}{2^{n/3}}$.

By assuming $q_m \leq 2^{2n/3}$, from each of the above cases we have

$$\Pr[\mathsf{B5}] \leq \frac{q_m \epsilon_2}{2^{n/3}}. \tag{17}$$

Bounding B6: Recall that, the event $\mathsf{B6}$ holds if $\exists i \in [q_m]$ such that $N_i = \mathsf{H}_{K_h}(M_i)$. For a fixed $i \in [q_m]$, the event holds with probability ϵ_1 due to the regular property of the hash function. Summing over all choices of i, we obtain the bound

$$\Pr[\mathsf{B6}] \leq q_m \epsilon_1. \tag{18}$$

Finally, by assuming $q_m \leq 2^{2n/3}$, Lemma 3 follows from Eqn. (21)-(18). □

6 1K-DWCDM : A Single Keyed DWCDM

Recall that, our proposed construction DWCDM is instantiated with a hash function and a block cipher where the hash key is independent to block cipher keys, leading to have a two-keyed (counting hash key separately from block cipher keys) nonce based MAC. In this section, we transform the DWCDM construction to a purely single keyed construction by setting the underlying hash key K_h to the encryption of $\mathbf{1}$ (i.e. $K_h := \mathsf{E}_K(\mathbf{1})$) and argue that the modified construction (that we call as 1K-DWCDM) is secure.

Now, we state and prove that 1K-DWCDM is secure up to $2^{2n/3}$ MAC queries and 2^n verification queries against all nonce respecting adversaries. We mainly focus on the nonce respecting security of the construction, as its nonce misuse security is very similar to that of DWCDM and hence we skip it.

Theorem 5. *Let $\mathcal{M}$ and $\mathcal{K}$ be finite and non-empty sets. Let $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $\mathsf{H} : \mathsf{E}_K(\mathbf{1}) \times \mathcal{M} \to \{0,1\}^n$ be an ϵ_2-AXU, ϵ_3-3-way regular and ϵ_1-regular hash function. Then, the MAC advantage of* 1K-DWCDM *is given by:*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{1K\text{-}DWCDM}[\mathsf{E},\mathsf{E}^{-1},\mathsf{H}]}(q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m + q_v, t') + \frac{3q_m}{2^{2n/3}} + \frac{q_m^2 \epsilon_2}{2^n} + \frac{q_v}{2^n - 1} + \max\{2q_v\epsilon_3, 2q_v\epsilon_2, q_v\epsilon_1\} + q_v\epsilon_1 + \frac{q_m}{2^n} + \frac{5q_m^3}{2^{2n}},$$

where $t' = O(t + (q_m + q_v)t_H)$, t_H being the time for computing hash function. Assuming ϵ_1, ϵ_2 and $\epsilon_3 \approx 2^{-n}$ and $q_m \leq 2^{2n/3}$, 1K-DWCDM$[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$ *construction is secured up to roughly $2^{2n/3}$ MAC and 2^n verification queries.*

Proof. The proof approach is similar to the one used in Theorem 2. Using standard argument, we can replace E_K and E_K^{-1} with n-bit uniform random permutation Π and its inverse Π^{-1}, denote the construction as $\mathsf{1K\text{-}DWCDM}^*[\Pi, \mathsf{E}^{-1}, \mathsf{H}]$ and bound the following:

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{1K\text{-}DWCDM}[\mathsf{E},\mathsf{E}^{-1},\mathsf{H}]}(q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m+q_v, t') + \underbrace{\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{1K\text{-}DWCDM}^*[\Pi,\Pi^{-1},\mathsf{H}]}(\mathsf{A})}_{\delta^*}. \tag{19}$$

Now, our goal is to prove the following:

$$\delta^* \leq \frac{3q_m}{2^{2n/3}} + \frac{q_m^2\epsilon_2}{2^n} + \max\{2q_v\epsilon_3, 2q_v\epsilon_2, q_v\epsilon_1\} + q_v\epsilon_1 + \frac{q_m}{2^n} + \frac{5q_m^3}{2^{2n}} + \frac{q_v}{2^n - 1}. \tag{20}$$

which we will do by applying H-Coefficient Technique in a similar way as we did in proving Theorem 2. For this, we first define the ideal oracle $(\mathsf{Rand}, \mathsf{Rej})$ which works as follows: for each MAC query (N, M), it samples the response T from $\{0,1\}^n$ uniformly at random and returns it to the distinguisher and for each verification query it returns $\perp$. For convenience, we slightly modify the experiment where we reveal to the distinguisher (after it made all it's queries and obtains corresponding responses but before it output it's decision) the hashing key K_h which is $\mathsf{E}_K(\mathbf{1})$, if we are in the real world, or a uniformly random dummy key K_h, sampled uniformly at random from $\{0,1\}^n$, if we are in the ideal world. All in all, the transcript of the attack is $\tau = (\tau_m, \tau_v, K_h)$ where τ_m and τ_v is the tuple of MAC and verification queries respectively.

Bad Transcript. The definition of bad transcript is similar to that of defined in section 5.2 and therefore, we have the following result:
Let X_{id} and Θ_b be defined as above. If $q_m \leq 2^{2n/3}$ and $q_v \leq 2^n$, then

$$\Pr[X_{\mathrm{id}} \in \Theta_b] \leq \frac{3q_m}{2^{2n/3}} + \frac{q_m^2\epsilon_2}{2^n} + \max\{2q_v\epsilon_3, 2q_v\epsilon_2, q_v\epsilon_1\} + q_v\epsilon_1 + \frac{q_m}{2^n}. \tag{21}$$

Analysis of Good Transcripts. Similar to Lemma 4, we prove that for any good transcript τ, realizing τ is almost as likely as real and in the ideal world. As the transcript τ is good, each sampled T_i value is non-zero. Since, in the ideal world the MAC oracle is perfectly random and the verification always rejects, one simply has

$$\mathsf{p}_{\mathrm{id}} := \Pr[X_{\mathrm{id}} = \tau] = \frac{1}{2^n} \cdot \frac{1}{(2^n - 1)^{q_m}} \tag{22}$$

Now, for the real interpolation probability, we have

$$\Pr[\Pi(N_i) \oplus \Pi(T_i) = \lambda_i, \forall i \in [q_m] \text{ and } \Pi(N'_a) \oplus \Pi(T'_a) \neq \lambda'_a, \forall a \in [q_v]].$$

Additionally, if the adversary makes any verification query (N'_a, M'_a, T'_a) with tag T'_a set to $\mathbf{1}$, then we need to ensure that

$$\Pi(N'_a) \neq \underbrace{\Pi(\mathbf{1}) \oplus N'_a \oplus \mathsf{H}_{\Pi(\mathbf{1})}(M'_a)}_{\lambda''_a}, \forall a \in [q_v]]. \tag{23}$$

Since, the hash key, i.e. $\Pi(\mathbf{1})$, is revealed to the adversary after the interaction is over, the right hand side of the non-Eqn. (23) becomes a constant, which makes it a uni-variate affine non-equation and then it is satisfied by condition (c) of Theorem 1. Therefore, we have

$$\begin{aligned}\mathsf{p}_{\mathrm{re}}(\tau) &= \frac{1}{2^n} \cdot \Pr[\Pi(N_i) \oplus \Pi(T_i) = \lambda_i, \forall i \in [q_m], \Pi(N'_a) \oplus \Pi(T'_a) \neq \lambda'_a, \\ &\qquad \Pi(N'_a) \neq \lambda''_a, \forall a \in [q_v]] \\ &\geq \frac{1}{2^n} \cdot \frac{1}{(2^n-1)^{q_m}} \cdot \left(1 - \frac{5q_m^3}{2^{2n}} - \frac{q_v}{2^n-1}\right) \qquad (24)\end{aligned}$$

The last inequality follows using similar to the proof of Lemma 4 and Eqn. (9). Finally, from Eqn. (22) and Eqn. (24), we compute the ratio as follows:

$$\frac{\mathsf{p}_{\mathrm{re}}(\tau)}{\mathsf{p}_{\mathrm{id}}(\tau)} \geq \left(1 - \frac{5q_m^3}{2^{2n}} - \frac{q_v}{2^n-1}\right). \qquad (25)$$

Eqn. (20) follows from Eqn. (21) and Eqn. (25). Finally, Theorem 5 follows from Eqn. (19) and Eqn. (20). □

7 Towards Higher Security of DWCDM

In this section, we briefly describe how to boost the security of DWCDM upto $(k-1)/k$-bit for a general k. The underlying construction remains as it is, however the nonce space is increased to $(k-1)n/k$-bits i.e. $\mathsf{DWCDM\_k}[\mathsf{E},\mathsf{H}](N,M) := \mathsf{E}_K^{-1}(\mathsf{E}_K(N) \oplus N \oplus \mathsf{H}_{K_h}(M))$ but here we consider $N = N^* \| 0^{n/k}$ where N^* is a $(k-1)n/k$ bit nonce. For this, we first state the following conjecture on Mirror theory, which is a generalized version of extended Mirror theorem as introduced in section. 3.2.

Conjecture 1 (Extended Mirror Theorem for $\xi_{\max} = k$). Let $(\mathcal{E}^{=} \cup \mathcal{E}^{\neq}, \phi', \Lambda')$ be a system of q many affine equations and v many affine non-equations associated with index mapping function ϕ' over $\mathrm{GF}(2^n)$ which are of the form $P_{\phi(n_i)} \oplus P_{\phi(t_i)} = \lambda_i$ for $i \in [q]$ and $P_{\phi(n_j)} \oplus P_{\phi(t_j)} \neq \lambda'_j (\neq \mathbf{0})$ for $j \in [q+1, q+v]$ over the set of α many unknown variables $\mathcal{P} = \{P_1, \ldots, P_\alpha\}$ such that P_a may be equals to some $P_{\phi(n_i)}$ or $P_{\phi(b_j)}$, where $a \in \{\phi(n_j), \phi(t_j)\}, j \in [q, q+v]$. Now, if

- (i) (ϕ', Λ') is good and
- (ii) $\xi_{\max} = k$

then the number of solutions for $\mathcal{P}$, denoted by h_β (where $\beta = \frac{kq}{k-1}$) such that $P_i \neq P_j$ for all distinct $i, j \in \{1, \ldots, \alpha\}$ is

$$h_\beta \geq \frac{(2^n)_\beta}{2^{nq}} \left(1 - O\left(\frac{q^k}{2^{(k-1)n}} + \frac{v}{2^n}\right)\right). \qquad (26)$$

Assuming this conjecture holds, we have the following result on the MAC advantage of DWCDM_k:

Theorem 6. *Let* E *be a block cipher and* H *be an* ϵ*-*j*-way regular hash function for all* $j \leq k$ *(e.g* PolyHash*). Then, the MAC advantage for any* (q_m, q_v, t) *nonce-respecting adversary against* $\mathsf{DWCDM\_k}$ *is given by,*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{DWCDM\_k}}(q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q_m + q_v, t') + O(q_m^k/2^{n(k-1)} + q_v.\epsilon),$$

where $t' = O(t + (q_m + q_v)t_H)$.

The proof will be similar to the proof of Theorem 2. We first define the transcript, associated MAC and the verification graph as before.

Now, we call a transcript $\tau = (\tau_m, \tau_v, K_h)$ to be bad if the associated MAC graph $G^{\mathtt{m}}_\tau$ and the Verification graph $G^{\mathtt{v}}_\tau$ satisfies the either of the following properties:

- $\mathsf{B1}'$: $G^{\mathtt{m}}_\tau$ has a component of size k or more.
- $\mathsf{B2}'$: $G^{\mathtt{m}}_\tau$ contains a valid cycle of length less than k.
- $\mathsf{B3}'$: $G^{\mathtt{v}}_\tau$ contains a valid cycle of length less than or equals to k that involves the verification query.

Moreover, τ is also said to be bad if it satisfies $\mathsf{B0}, \mathsf{B4}, \mathsf{B5}, \mathsf{B6}$ (as defined in Definition 4).
Here we will mainly consider bounding B1', B2' and B3', as the remaining ones are already done. Here we provide a sketch for bounding each of this event:

Bounding B1'. Event B1' occurs if there exists a component of size at least k in $G^{\mathtt{m}}_\tau$. This essentially implies there is a chain of $(k-1)$ edges. Let there are c_1 number of edges are of the form $T_i = N_j$ with $i < j$. Here we claim that

$$\Pr[\mathsf{B1}'] \leq q_m.\left(\frac{q_m}{2^n}\right)^{k-c_1}.\left(\frac{1}{2^{k/n}}\right)^{c_1}. \tag{27}$$

As $k \geq 4$, the above bound is $O(q_m^k/2^{n(k-1)})$.

Bounding B2'. Event B2' occurs if there exists a cycle of size less than k in $G^{\mathtt{m}}_\tau$. Let us bound a cycle of length $c < (k-1)$. Again, assume there are c_1 number of edges of the form $T_i = N_j$ with $i < j$. Using similar argument as above,

$$\Pr[\mathsf{B2}'] \leq \left(\frac{q_m}{2^n}\right)^{c-c_1}.\left(\frac{1}{2^{k/n}}\right)^{c_1}. \tag{28}$$

It is easy to see that for any c, the above bound is $O(q_m/2^n)$.

Bounding B3'. Event B3' occurs if there exists a cycle of size less than or equals to k in $G^{\mathtt{v}}_\tau$. Extending similar arguments used in lemma 3 to bound the event B3, one can show that if H is ϵ j-way regular for all $j \leq k$ then

$$\Pr[\mathsf{B3}'] \approx O\left(\frac{q_v.\epsilon.q_m^c}{2^{nc}}\right), \tag{29}$$

for some $c \geq 0$.

Combining everything together, we can bound

$$\begin{aligned}\Pr[\mathsf{B}] &\leq \Pr[\mathsf{B0}] + \Pr[\mathsf{B1'}] + \Pr[\mathsf{B2'}] + \Pr[\mathsf{B3'}] + \Pr[\mathsf{B4}] + \Pr[\mathsf{B5}] + \Pr[\mathsf{B6}] \\ &\approx O(q_m^k/2^{n(k-1)} + q_v.\epsilon)\end{aligned}$$

Next, we fix a good transcript τ. Now, to obtain the lower bound of the probability of getting τ in real world, we need a lower bound on the probability of the number of solutions to a system of q_m many equations and q_v many non-equations. Again, we can do that using an extended Mirror theory result with maximal block size $\xi_{\max} = k$. From Conjecture 1, we have

$$\mathsf{p}_{\mathrm{re}}(\tau) \geq \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{2^{nq_m}} \cdot \left(1 - O\left(\frac{q_m^k}{2^{(k-1)n}} + \frac{q_v}{2^n}\right)\right), \tag{30}$$

The theorem follows by applying Patarin's H-Coefficient Technique. □

Remark 2. We would like to clarify that increasing the nonce space does not have any relation with the increase in security. We have restricted the nonce space of DWCDM to $2n/3$-bit (note that this is minimum as we must allow $2^{2n/3}$ many MAC queries with distinct nonces) purely because of the simplicity of the extended mirror theory analysis. One can of course increase the nonce space to $(k-1)n/k$-bit for any $k \leq n$, but that increases the block maximality ($\xi_{\max}$) to k and hence the analysis of the extended mirror theory would become tedious and involved..

Acknowledgments

Initial part of this work was done in NTT Lab, Japan when Avijit Dutta was visiting there. Mridul Nandi is supported by R.C.Bose Centre for Cryptology and Security. The authors would like to thank all the anonymous reviwers of CRYPTO 2018 for their invaluable comments and suggestions that help to improve the overall quality of the paper.

References

1. Ray Beaulieu, Douglas Shors, Jason Smith, Stefan Treatman-Clark, Bryan Weeks, and Louis Wingers. The simon and speck families of lightweight block ciphers. Cryptology ePrint Archive, Report 2013/404, 2013. http://eprint.iacr.org/2013/404.
2. Christof Beierle, Jérémy Jean, Stefan Kölbl, Gregor Leander, Amir Moradi, Thomas Peyrin, Yu Sasaki, Pascal Sasdrich, and Siang Meng Sim. The SKINNY family of block ciphers and its low-latency variant MANTIS. In *CRYPTO 2016, Proceedings, Part II*, pages 123–153, 2016.
3. M. Bellare and R. Impagliazzo. A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to prp to prf conversion. Cryptology ePrint Archive, Report 1999/024, 1999. http://eprint.iacr.org/1999/024.

4. Mihir Bellare, Ran Canetti, and Hugo Krawczyk. Keying hash functions for message authentication. In *Advances in Cryptology - CRYPTO '96*, pages 1–15, 1996.
5. Mihir Bellare, Oded Goldreich, and Anton Mityagin. The power of verification queries in message authentication and authenticated encryption. *IACR Cryptology ePrint Archive*, 2004:309, 2004.
6. Mihir Bellare, Joe Kilian, and Phillip Rogaway. The security of cipher block chaining. In *CRYPTO '94*, pages 341–358, 1994.
7. Mihir Bellare, Ted Krovetz, and Phillip Rogaway. Luby-rackoff backwards: Increasing security by making block ciphers non-invertible. In *EUROCRYPT '98, Proceeding.*, pages 266–280, 1998.
8. Srimanta Bhattacharya and Mridul Nandi. Full indifferentiable security of the xor of two or more random permutations using the \chi ˆ2 method. In *EUROCRYPT 2018 Proceedings, Part I*, pages 387–412, 2018.
9. Srimanta Bhattacharya and Mridul Nandi. Revisiting variable output length XOR pseudorandom function. *IACR Trans. Symmetric Cryptol.*, 2018(1):314–335, 2018.
10. Srimanta Bhattacharya and Mridul Nandi. Revisiting variable output length XOR pseudorandom function. *IACR Trans. Symmetric Cryptol.*, 2018(1):314–335, 2018.
11. Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and C. Vikkelsoe. PRESENT: an ultra-lightweight block cipher. In *CHES 2007, Proceedings*, pages 450–466, 2007.
12. Christophe De Cannière, Orr Dunkelman, and Miroslav Knezevic. KATAN and KTANTAN - A family of small and efficient hardware-oriented block ciphers. In *CHES 2009, Proceedings*, pages 272–288, 2009.
13. Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the two-round even-mansour cipher. In *CRYPTO 2014. Proceedings, Part I*, pages 39–56, 2014.
14. Benoit Cogliati, Rodolphe Lampe, and Jacques Patarin. The indistinguishability of the XOR of k permutations. In *FSE 2014. Revised Selected Papers*, pages 285–302, 2014.
15. Benoît Cogliati and Yannick Seurin. EWCDM: an efficient, beyond-birthday secure, nonce-misuse resistant MAC. In *CRYPTO 2016, Proceedings, Part I*, pages 121–149, 2016.
16. Benoît Cogliati and Yannick Seurin. Analysis of the single-permutation encrypted davies-meyer construction. *To be appeared in Des. Codes Cryptography*, 2018, 2018.
17. Joan Daemen and Vincent Rijmen. Rijndael for AES. In *AES Candidate Conference*, pages 343–348, 2000.
18. Wei Dai, Viet Tung Hoang, and Stefano Tessaro. Information-theoretic indistinguishability via the chi-squared method. In *Advances in Cryptology - CRYPTO 2017, Proceedings, Part III*, pages 497–523, 2017.
19. Nilanjan Datta, Avijit Dutta, Mridul Nandi, Goutam Paul, and Liting Zhang. Single key variant of pmac_plus. *IACR Trans. Symmetric Cryptol.*, 2017(4):268–305, 2017.
20. Avijit Dutta, Ashwin Jha, and Mridul Nandi. Tight security analysis of ehtm MAC. *IACR Trans. Symmetric Cryptol.*, 2017(3):130–150, 2017.
21. Jian Guo, Thomas Peyrin, Axel Poschmann, and Matthew J. B. Robshaw. The LED block cipher. *IACR Cryptology ePrint Archive*, 2012:600, 2012.
22. Tetsu Iwata. New blockcipher modes of operation with beyond the birthday bound security. In *FSE 2006, Revised Selected Papers*, pages 310–327, 2006.
23. Tetsu Iwata, Bart Mennink, and Damian Vizár. CENC is optimally secure. *IACR Cryptology ePrint Archive*, 2016:1087, 2016.

24. Stefan Lucks. The sum of prps is a secure PRF. In *EUROCRYPT 2000*, pages 470–484, 2000.
25. Bart Mennink and Samuel Neves. Encrypted davies-meyer and its dual: Towards optimal security using mirror theory. In *CRYPTO 2017. Proceedings, Part III*, pages 556–583, 2017.
26. Kazuhiko Minematsu and Tetsu Iwata. Building blockcipher from tweakable blockcipher: Extending FSE 2009 proposal. In *IMACC 2011. Proceedings*, pages 391–412, 2011.
27. Yusuke Naito. Blockcipher-based macs: Beyond the birthday bound without message length. In *ASIACRYPT 2017. Proceedings, Part III*, pages 446–470, 2017.
28. NIST. Recommendation for block cipher modes of operation: The CMAC mode for authentication. *SP 800-38B*, 2005.
29. Jacques Patarin. A proof of security in o(2n) for the xor of two random permutations. In *ICITS 2008, Proceedings*, pages 232–248, 2008.
30. Jacques Patarin. The "Coefficients H" Technique. In *Selected Areas in Cryptography, SAC*, pages 328–345, 2008.
31. Jacques Patarin. Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. *IACR Cryptology ePrint Archive*, 2010:287, 2010.
32. Jacques Patarin. Security in $o(2^n)$ for the xor of two random permutations - proof with the standard H technique. *IACR Cryptology ePrint Archive*, 2013:368, 2013.
33. Jacques Patarin. Mirror theory and cryptography. *Appl. Algebra Eng. Commun. Comput.*, 28(4):321–338, 2017.
34. Kan Yasuda. A new variant of PMAC: beyond the birthday bound. In *CRYPTO 2011*, pages 596–609, 2011.

A Proof of Theorem 3

Proving the MAC security in nonce misuse setting, PRF security of the construction implies its MAC security and hence we view the construction as a keyed function with domain $\mathcal{N} \times \mathcal{M}$ and study the PRF security of it. Our main security result about $\mathsf{DWCDM}[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$ against nonce misuse adversary is as follows.

Lemma 5. *Let $\mathcal{M}, \mathcal{K}$ and $\mathcal{K}_h$ be finite and non-empty sets. Let $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $\mathsf{H} : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ϵ_1-regular and ϵ_2-AXU hash function. Then, the PRF advantage of $\mathsf{DWCDM}[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$ is given by,*

$$\mathbf{Adv}^{\mathrm{PRF}}_{\mathsf{DWCDM}[\mathsf{E},\mathsf{E}^{-1},\mathsf{H}]}(q,t) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(q,t') + q^2\epsilon_2 + q\epsilon_1 + \frac{4q^2}{2^n} + \frac{q}{2^n},$$

where $t' = O(t + qt_H)$, t_H be the time for computing hash function.

Theorem 3 follows from Lemma 5 and the folklore result that a secure PRF is a secure MAC due to Bellare et al. [5]:

A.1 Proof of Lemma 5

As before, we fix a non-repeating query making (q, t) distinguisher A against the PRF security of $\mathsf{DWCDM}[\mathsf{E}, \mathsf{E}^{-1}, \mathsf{H}]$. As the first step of the proof, we replace E_K and E_K^{-1} with a n-bit uniform random permutation Π and its inverse Π^{-1} respectively at the cost of $\mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(\mathsf{A}')$ and denote the resulting construction as $\mathsf{DWCDM}^*[\Pi, \Pi^{-1}\mathsf{H}]$, where A' is an adversary against the SPRP security of E, making at most q queries and running time is at most $t + qt_H$ such that

$$\mathbf{Adv}^{\mathrm{PRF}}_{\mathsf{DWCDM}[\mathsf{E},\mathsf{E}^{-1},\mathsf{H}]}(\mathsf{A}) \leq \mathbf{Adv}^{\mathrm{SPRP}}_{\mathsf{E}}(\mathsf{A}') + \mathbf{Adv}^{\mathrm{PRF}}_{\mathsf{EWCDM}^*[\Pi,\Pi^{-1},\mathsf{H}]}(\mathsf{A}) \tag{31}$$

It remains to upper bound the PRF advantage of A against $\mathsf{DWCDM}^*[\Pi, \Pi^{-1}, \mathsf{H}]$. For this we apply H-Coefficient Technique as follows: Adversary A interacts (a) either with the real oracle $\mathsf{DWCDM}^*[\Pi, \Pi^{-1}, \mathsf{H}]$ for a random permutation Π, its inverse Π^{-1} and a random hashing key K_h or (b) with the ideal oracle in which it receives uniformly and independent responses against each distinct query it makes. Note that, the adversary is allowed to make distinct query with same nonce.

After the interaction is over (but before A output its decision bit), we release the actual hash key K_h to the adversary A, if A is in the real world, or a random dummy hash key K_h if A is in the ideal world.

A.1.1 Attack Transcript and Transcript Graph. Let the transcript of A be $\tau = (\tau_m, K_h)$ where $\tau_m := \big((N_1, M_1, T_1), \ldots, (N_q, M_q, T_q)\big)$ be the list of queries and reponses of A. Θ, X_{re} and X_{id} respectively denotes the set of all attainable transcripts, the probability distribution of transcript τ induced by the real world and ideal world.

Similar to proof of Theorem 2, we define a transcript graph G_τ corresponding to a transcript τ as follows:

$$G_\tau = ([q], E) \text{ where } E = \{(i, j) \in [q]\times[q] : N_i = N_j \vee N_i = T_j \vee N_j = T_i \vee T_i = T_j\}.$$

For the sake of convenience of the proof, we denote the edge (i, j) as a dotted line when $T_i = T_j$ and as a dashed line when $N_i = N_j$. Else, we denote it as a continuous line. Thus, the edge set of G_τ consists of three different types of edges as depicted in Fig. 5.1 (a), (b) and (c).

A.1.2 Definition and Probability of Bad Transcripts.

Definition 5. *An attainable transcript $\tau = (\tau_m, K_h)$ is* bad *if*

- B1 : *there exists two distinct queries (N_i, M_i, T_i) and (N_j, M_j, T_j) such that $T_i = T_j$.*
- B2 : *there exists two distinct queries (N_i, M_i, T_i) and (N_j, M_j, T_j) such that $N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j), N_i = N_j$.*
- B3 : *$\exists i \in [q]$ such that $\mathsf{H}_{K_h}(M_i) = N_i$.*

- $\mathsf{B4}$: $\exists i \in [q]$ *such that* $N_i = T_i$.
- $\mathsf{B5}$: *if there is a cycle of length at least* 2 *in* G_τ.

Let $\Theta_b \subseteq \Theta$ denotes the set of all attainable bad transcripts. In the following lemma, we upper bound the probability of realizing a bad transcript in ideal world.

Lemma 6. *Let* X_{id} *and* Θ_b *be defined as above. Then we have,*

$$\Pr[X_{\text{id}} \in \Theta_b] \leq \epsilon_{\text{bad}} = q^2\epsilon_2 + q\epsilon_1 + \frac{3q^2}{2^n} + \frac{q}{2^n}$$

Proof. Let us denote $\mathsf{B} := \mathsf{B1} \vee \mathsf{B2} \vee \mathsf{B3} \vee \mathsf{B4} \vee \mathsf{B5}$. In order to bound $\Pr[X_{\text{id}} \in \Theta_b]$, it is enough to bound $\Pr[\mathsf{B}]$. Therefore, we have

$$\Pr[\mathsf{B}] \leq \Pr[\mathsf{B1}] + \Pr[\mathsf{B2}] + \Pr[\mathsf{B3}] + \Pr[\mathsf{B4}] + \Pr[\mathsf{B5} \mid \overline{\mathsf{B1}}] \tag{32}$$

In the following, we bound the probabilities of all the bad events individually.

Bounding B1: Fix two distinct query indices i and j such that $T_i = T_j$. For two fixed distinct indices $i, j \in [q]$, the event holds with probability 2^{-n} as T_i's are sampled uniformly and independent to all previously sampled random variables in the ideal world. Summing over all possible choices of i and j, we obtain the bound

$$\Pr[\mathsf{B1}] \leq \frac{q^2}{2^n}. \tag{33}$$

Bounding B2: Fix two distinct query indices i and j such that $N_i \oplus \mathsf{H}_{K_h}(M_i) = N_j \oplus \mathsf{H}_{K_h}(M_j), N_i = N_j$. For two fixed distinct indices $i, j \in [q]$, the event holds with probability ϵ_2 due to the randomness of the hash key K_h.as the hash key is sampled uniformly and independent to all previously sampled random variables in the ideal world. Summing over all possible choices of i and j, we obtain the bound

$$\Pr[\mathsf{B2}] \leq q^2\epsilon_2. \tag{34}$$

Bounding B3 and B4: Observe that, $\mathsf{B3}$ and $\mathsf{B4}$ have already been defined in Definition 4. Therefore, following Lemma 3, we have

$$\Pr[\mathsf{B3}] \leq q\epsilon_1, \quad \Pr[\mathsf{B4}] \leq \frac{q}{2^n} \tag{35}$$

where the probability of the former event is bounded by the regularity assumption on the hash function whereas the later one is bounded by the randomness of T_i.

Bounding B5 | $\overline{\mathsf{B1}}$: We bound the probability of formation of a cycle in the transcript graph G_τ when all T's are distinct. Let us consider such a cycle of length p and we denote the corresponding event by $\mathsf{Cyl_p}$. Note that, since T's are all distinct, there cannot be any dotted edge in the cycle. Moreover, there

cannot be any dashed edge in the cycle as that would leads to have an invalid cycle.

Example. Let us consider the following set of four equations:

$$\begin{cases} \Pi(N_1) \oplus \Pi(T_1) = N_1 \oplus \mathsf{H}_{K_h}(M_1) \\ \Pi(N_2) \oplus \Pi(T_2) = N_2 \oplus \mathsf{H}_{K_h}(M_2) \\ \Pi(N_3) \oplus \Pi(T_3) = N_3 \oplus \mathsf{H}_{K_h}(M_3) \\ \Pi(N_4) \oplus \Pi(T_4) = N_4 \oplus \mathsf{H}_{K_h}(M_4) \end{cases}$$

In the above set of equations all T's are distinct. Now, consider a cycle C of length 4 where there is only one dashed edge (wlog of we assume the dashed edge is in between 1 and 2) and all remaining three edges are continuous. Then, C imposes the following equality pattern on (N, T):

$$N_1 = N_2, N_2 = T_3, N_3 = T_4, N_4 = T_1.$$

These set of constraints eventually leads us to have $\Pi(T_3) = \bigoplus_{i=1}^{4}(N_i \oplus \mathsf{H}_{K_h}(M_i))$. Hence the cycle is invalid. It is simple to observe that existence of any dashed edge in any cycle of G_τ makes it invalid when all T's are distinct. Hence, if the cycle has to be valid, then there must be only continuous edges in the cycle.

Let us consider such a valid cycle $C = (i_1, i_2, \ldots, i_p)$ of length p. Without loss of generality we may assume $i_1 < i_2 < \ldots < i_p$. Therefore, from Lemma 2, the probability of each of the edges (i_k, i_{k+1}) for all $k = 1, \ldots, p-1$ is 2^{-n} and as $i_p > i_1$, probability of the edge (i_p, i_1) is $2^{-n/3}$. Therefore, the for a fixed choice of indices $i_1, \ldots, i_p$, the probability of the cycle becomes $\frac{1}{2^{n(p-1)+n/3}}$. Hence, by summing over all possible choices of $i_1, \ldots, i_p$, we have

$$\Pr[\mathsf{Cyl_p}] \leq \frac{q^p}{2^{n(p-1)+n/3}}.$$

Therefore,

$$\Pr[\mathsf{B5} \mid \overline{\mathsf{B1}}] \leq \sum_{p=2}^{\infty} \Pr[\mathsf{Cyl_p}] = \frac{q^2}{2^{4n/3}}\left(1 + \frac{q}{2^n} + \frac{q^2}{2^{2n}} + \ldots\right) = \frac{q^2}{2^{4n/3}} \cdot \frac{1}{(1 - \frac{q}{2^n})}$$

As we assume, $q \leq 2^n/2$, we have

$$\Pr[\mathsf{B5} \mid \overline{\mathsf{B1}}] \leq \frac{2q^2}{2^{4n/3}} \leq \frac{2q^2}{2^n} \tag{36}$$

Finally, Lemma 6 follows from Eqn. (32), Eqn. (33), Eqn. (34), Eqn. (35) and Eqn. (36). □

A.1.3 Analysis of Good Transcripts. Now, we analyze good transcripts in the following lemma:

Lemma 7. *Let $\tau = (\tau_m, K_h)$ be a good transcript. Then*

$$\frac{\mathsf{p}_{\mathrm{re}}(\tau)}{\mathsf{p}_{\mathrm{id}}(\tau)} := \frac{\Pr[X_{\mathrm{re}} = \tau]}{\Pr[X_{\mathrm{id}} = \tau]} \geq (1 - \epsilon_{\mathrm{ratio}}) = (1 - \frac{q^2}{2^n}).$$

Proof. Let $\tau = (\tau_m, K_h)$ be a good transcript. Then one has

$$\Pr[X_{\mathrm{id}} = \tau] = \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{2^{nq}} \tag{37}$$

as in the ideal world, the oracle is perfectly random and the hash key K_h is chosen uniformly at random and independently from the query transcript.

Since, τ is good, the following set of equations

$$\mathcal{E} = \begin{cases} \Pi(N_1) \oplus \Pi(T_1) = N_1 \oplus \mathsf{H}_{K_h}(M_1) \\ \Pi(N_2) \oplus \Pi(T_2) = N_2 \oplus \mathsf{H}_{K_h}(M_2) \\ \vdots \\ \Pi(N_q) \oplus \Pi(T_q) = N_q \oplus \mathsf{H}_{K_h}(M_q) \end{cases}$$

is circle free and contains no collision in the solution within a same block. Let us assume that the maximal block size of $\mathcal{E}$ is $\xi_{\max}$. Then due to Theorem 3 of [31], which is known as "*Theorem $P_i \oplus P_j$ when $\xi_{\max}\alpha \ll 2^n$*", we have

$$\Pr[X_{\mathrm{re}} = \tau] = \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{2^{nq}}(1 - \frac{q\xi_{\max}}{2^n}) \tag{38}$$

Computing the ratio of Eqn. (38) and Eqn. (37), we have

$$\frac{\mathsf{p}_{\mathrm{re}}(\tau)}{\mathsf{p}_{\mathrm{id}}(\tau)} \geq (1 - \frac{q\xi_{\max}}{2^n}) \geq (1 - \frac{q^2}{2^n}).$$

where the last inequality follows as $\xi_{\max} \leq q$. □

Finally, Lemma 5 follows from Lemma 6, Lemma 7 and Lemma 1.