Beyond Birthday Bound Secure MAC in Faulty Nonce Model [*]

Avijit Dutta, Mridul Nandi and Suprita Talnikar

Indian Statistical Institute, Kolkata.
avirocks.dutta13@gmail.com, mridul.nandi@gmail.com, suprita45@gmail.com

Abstract. Encrypt-then-MAC (EtM) is a popular mode for authenticated encryption (AE). Unfortunately, almost all designs following the EtM paradigm, including the AE suites for TLS, are vulnerable against nonce misuse. A single repetition of the nonce value reveals the hash key, leading to a universal forgery attack. There are only two authenticated encryption schemes following the EtM paradigm which can resist nonce misuse attacks, the GCM-RUP (CRYPTO-17) and the $\mathsf{GCM}/2^{+}$ (INSCRYPT-12). However, they are secure only up to the birthday bound in the nonce respecting setting, resulting in a restriction on the data limit for a single key. In this paper we show that nEHtM, a nonce-based variant of EHtM (FSE-10) constructed using a block cipher, has a beyond birthday bound (BBB) unforgeable security that gracefully degrades under nonce misuse. We combine nEHtM with the CENC (FSE-06) mode of encryption using the EtM paradigm to realize a nonce-based AE, CWC+. CWC+ is very close (requiring only a few more xor operations) to the CWC AE scheme (FSE-04) and it not only provides BBB security but also gracefully degrading security on nonce misuse.

Keywords: Graceful Security, Faulty Nonce, Mirror Theory, Extended Mirror Theory, Expectation Method, CWC, GCM.

1 Introduction

An authenticated encryption (AE) mode is a cryptographic scheme that guarantees the privacy and authenticity of a message concurrently. Authenticated encryption has received much attention from the cryptographic community mostly due to its application to TLS and many other protocols. The ongoing CAESAR competition [1] which aims to identify a portfolio of authenticated encryption schemes has drafted three use cases, namely *lightweight*, *high-performance*, and *defense-in-depth.* The competition considers GCM [25] as the baseline algorithm as it is widely adopted (e.g. in TLS 1.2 and in its variant RGCM [6], which shall soon be considered in TLS 1.3 [11]) and standardized. ChaCha20+Poly1305 [7] is a popular alternative for settings where AES-NI is not implemented.

[*] This is the full version of the article accepted in IACR-EUROCRYPT 2019.

Encrypt-then-MAC. Both ChaCha20+Poly1305 and GCM follow the Encrypt-then-MAC (EtM) paradigm [5]. Some other popular AE designs following the same paradigm are CWC [23], GCM/2^+ [3], CHM [21], CIP [22], GCM-RUP [4], OGCM1 [40], OGCM2 [40] etc. The authenticated encryption of this paradigm is described as follows. Let $\mathcal{E}$ be a nonce-based encryption scheme and $\mathcal{I}$ be a message authentication code. Given a nonce N, a message M and an associated data A, the ciphertext $C = \mathcal{E}^N(M)$ is first computed, which is then used to compute the tag $T = \mathcal{I}(N, A, C)$. All the aforementioned algorithms can be described by an encryption $\mathcal{E}^N$ (involving stream cipher encryptions) and a MAC $\mathcal{I}$ (all constructions are algebraic hash function-based and most of them uses Wegman-Carter MAC (WC) [39]). EtM is a popular design paradigm due to its generic security guarantee. Authors of [12] showed that (stating informally) if $\mathcal{E}$ is a secure symmetric encryption scheme and $\mathcal{I}$ is a secure MAC family then this method of implementing EtM results in secure channels. This has later also been analyzed by [5, 30].

1.1 Nonce Misuse Resistance Security

As shown in Joux's forbidden attack [2], GCM turns out to leak the hash key whenever an encryption query with a repeated nonce is executed. A similar forgery attack can be applied against all aforementioned AE, except GCM-RUP and GCM/2^+, as they use some variants of WC MAC. GCM-RUP resists this attack as it uses the XEX [37] construction to define the tag. The tag of XEX for a data D is computed as $\mathsf{E}_K(\mathsf{H}_{K_h}(D) \oplus N) \oplus \mathsf{H}_{K_h}(D)$. However, in nonce-respecting settings it gives up to the birthday bound security due to the following attack.

- After making $2^{n/2}$ nonce-respecting queries, we expect a collision amongst the values of $\mathsf{H}_{K_h}(D) \oplus N$, where n is the block size of the underlying block cipher. This can be detected through a collision amongst the values of $N \oplus T$, where T denotes the tag. Whenever this collision happens, one knows the difference between the hash outputs, which eventually leaks the hash key.

GCM/2^+ resists the attack as it uses the Encrypted Wegman-Carter-Shoup (EWCS) [13] construction to define the tag. The tag of EWCS for a data D is computed as $\mathsf{E}_{K_2}(\mathsf{E}_{K_1}(N) \oplus \mathsf{H}_{K_h}(D))$. However, in nonce-respecting settings it gives up to the birthday bound PRF security as an adversary makes $2^{n/2}$ nonce respecting queries with the same message and observes no collision in the tag.

In some contexts, it becomes challenging to maintain the uniqueness of the nonce, for example on implementations in a stateless device or in cases where the nonce is chosen randomly from a small set. Moreover, due to the faulty implementation of the cipher or occurrence of some fault (for example, if the nonce gets reset), the nonce may repeat. After making an internet-wide scan, Böck et al. [9] found 184 devices that used a duplicate nonce. Thus a construction which provides security against nonce misuse is desirable.

1.2 Beyond Birthday Bound Security with Graceful Degradation

Achieving beyond birthday security would provide a larger data limit for a single key. GCM-RUP can be proven to have at most $\ell q_m^2/2^n$ forging advantage (in the nonce-respecting model), where q_m is the number of encryption queries and ℓ is the maximum number of data blocks a message and an associated data can possess. For example, the GCM-RUP based on AES which can process a data of size at most $\ell = 2^{32}$ blocks should have a data limit $q_m \leq 2^{32}$ so as to allow an advantage of at most 2^{-32}, a tolerance level much smaller than that provided by beyond birthday security. To achieve security beyond the birthday bound in a nonce respecting setting, *Encrypted Wegman-Carter with Davies-Meyer* [13] (or EWCDM) and *Decrypted Wegman-Carter with Davies-Meyer* [16] (or DWCDM) have been proposed. However, these constructions provide birthday bound security with only a single misuse of nonce. In other words, they do not provide graceful degradation of security in nonce misuse settings. There are other known constructions such as *Dual Encrypted Wegman-Carter with Davies-Meyer* (or EWCDMD) [26, 31], *Encrypted Wegman-Carter-Shoup* [13] (or EWCS) and single hash-key variants of CLRW2 [24]. However, these constructions provide birthday bound PRF security in nonce-respecting settings. Note that the PRF security of the MAC contributes to the privacy of the encryption of EtM constructions.

Goal of the paper. The main goal of this paper is to find *an efficient MAC which is BBB (beyond birthday bound) secure both as a PRF and a MAC*. Moreover, it should provide *graceful security degradation in a nonce-misuse setting*. It must be mentioned here that there are some deterministic MAC constructions (not requiring any nonce) that provide BBB security. These mainly follow a double-block hash-then-sum approach [14, 15] and hence require the computation of two blocks of algebraic hashes (or one pass of block cipher or tweakable block cipher executions). However, a single-block hash (which would be definitely faster than two blocks of hash and require a smaller hash-key size) would be a better option. So, this paper focuses on getting a design based on a single-block algebraic hash (e.g., a single-call of the polynomial hash [29]).

Graceful Degradation of Security on Nonce Misuse. The most popular metric to measure nonce misuse is the maximum number of multicollisions in nonce values amongst all queries [36]. To the best of our knowledge, none of the existing block cipher-based nonce-based MACs adhere to this notion with BBB security guarantee. We have also explored many other variants of MAC constructions using at most two block cipher calls and a single hash function call. Unfortunately, we found that none of them give beyond birthday bound security in terms of multicollision nonce misuse, even with multicollisions of size 2.

In this paper we instead consider another natural definition of nonce misuse, called the number of faulty nonces. An authentication query is said to be a *faulty query* if there exists a previous MAC query such that their corresponding nonces match. The nonce in a faulty query is called a *faulty nonce*. The notion of a faulty nonce is weaker than multicollision. When a counter is implemented in an aperiodic manner (e.g., timely nonce [9] used in TLS 1.2), a simple reset does not give a large number of faulty nonces; there are easy countermeasures to prevent a large number of faulty nonce encryptions.

1.3 Our Contribution

Our contribution in this paper is threefold which we outline as follows:

1. Multicollision on Universal Hash. We study the probability of occurrence of multicollisions in a universal hash function. In particular, we have shown that the probability of obtaining a $(\xi + 1)$-multicollision tuple amongst q inputs is at most $q^2\epsilon/\xi$ (see Sect. 5). This is clearly an improved bound as compared to a straightforward application of the union bound. We believe that this problem can have independent interest in the cryptographic community and can be used to get improved bounds for other constructions also.

2. BBB Secure MAC with Graceful Security. In [28], a probabilistic MAC EHtM has been analyzed and shown to have roughly $3n/4$-bit MAC security which is also tight [18]. This paper analyzes a construction, which shall be denoted as nEHtM, where (1) the random salt is replaced by the nonce and (2) the two independent pseudorandom functions are replaced by a single-keyed block cipher. Given a data D and a nonce N the tag is computed as follows (see Fig. 1.1(b)):

$$\mathsf{nEHtM}_{K,K_h}(N, D) \stackrel{\Delta}{=} \mathsf{E}_K(0\|N) \oplus \mathsf{E}_K(1\|\mathsf{H}_{K_h}(D) \oplus N).$$

We have shown that nEHtM is secure roughly up to $2^{2n/3}$ authentication queries and 2^n verification queries in the nonce-respecting setting. Moreover, this security degrades in a graceful manner on introduction of faults in the nonce. The

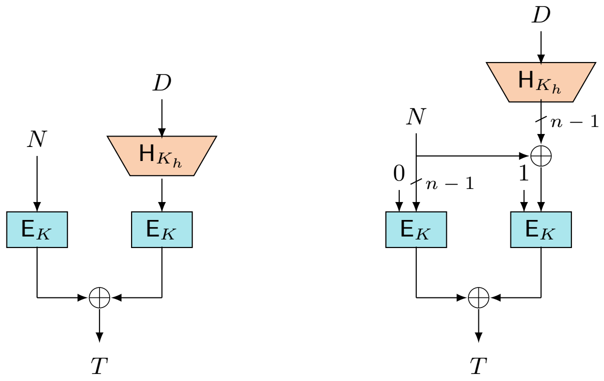

Fig. 1.1. (a) On the left is the CWC MAC (MAC algorithm used in CWC); (b) on the right is the domain separation variant of *nonce-based Enhanced Hash-then-Mask.*

unforgeability of this construction shall be shown through an extended distinguishing game. We apply the expectation method (as it shall later be shown to give a better bound than the coefficients-H technique) to bound the distinguishing advantage of two worlds. In the ideal world, once we realize the random tags T_i, we need to sample the hash key. This would determine all inputs of the underlying block cipher. The equality patterns amongst the nonce values are deterministic and we bound the number of faulty nonces by a parameter μ. However, the equality patterns among other inputs of the form $X \overset{\Delta}{=} \mathsf{H}_{K_h}(D) \oplus N$ are probabilistic due to randomness of the hash key. As there may not be sufficient entropy in the hash-key (which could be n-bit for polynomial hash), the number of multicollisions amongst the values of X may not be easy to compute. We have tackled this problem using the multicollision result (as stated in the first contribution) of the underlying hash function.

After we limit the multicollisions in the values of both X and N, we shall be in a position to apply mirror theory to show a beyond birthday bound security on the distinguishing advantage of nEHtM. Note that mirror theory cannot give a beyond birthday bound security without restricting the number of multicollisions.

It must be noted here that nEHtM (like all other candidates) is not secure beyond the birthday bound under the notion of multicollision nonce misuse security and the corresponding attack is discussed in Supplementary Sect. A.

3. Application to a CWC-like AE Construction. We propose CWC+, which is an instance of the EtM composition based on the CENC type encryption with maximum width parameter and the nEHtM MAC. Moreover, we apply an appropriate domain separation to make it a single-keyed construction (even the hash key is generated from the the block cipher). The construction is a very close variant of CWC as it requires a few additional xor computations, without requiring any extra calls to the block cipher. Furthermore, CWC+ gives both (1) BBB security and (2) graceful security degradation in the faulty nonce misuse model. In particular, we have the following forging advantage of CWC+:

$$\mathsf{Auth}[\mathsf{CWC+}] = \frac{105\sigma^3\ell}{2^{2n}} + \frac{6\sigma\ell}{2^n} + \frac{2q_d}{2^\rho} + \frac{2q_d\ell}{2^n} + \frac{(2q_e + q_d)2\ell\mu}{2^n} + \left(\frac{5\sigma\ell\mu}{2^n}\right)^2,$$

where q_e and q_d is the number of encryption and decryption queries, ρ is the tag size, ℓ is the maximum number of message blocks queried including the associated data blocks, σ is the total number of message blocks queried and μ is the total number of faulty queries. Moreover, the security of CWC+ gracefully drops to birthday bound when $\ell\mu$ is about $2^{n/2}$. However, when $\ell \leq 2^{n/4}$, then the security bound of CWC+ caps at roughly $2^{7n/12}$, which is strictly greater than the birthday bound. A better bound can be obtained if we assume some restrictions over all the message lengths.

(3) Another notable feature of CWC+ is that the scheme remains secure even with short tag lengths. In GCM, if the tag length is only 32 bits, then an adversary forges the construction with just 1024 verification attempts by querying with a single message consisting of 2^{22} blocks. However, for the same tag size,

authenticity advantage of $\mathsf{CWC+}$ is 2^{-21} when adversary forges the construction with 1024 verification attempts.

2 Preliminaries

Basic Notations: For a set $\mathcal{X}$, $X \leftarrow_{\$} \mathcal{X}$ denotes that X is sampled uniformly at random from $\mathcal{X}$ and is independent to all other random variables defined so far. $\{0,1\}^n$ denotes the set of all binary strings of length n and $\{0,1\}^*$ denotes the set of all binary strings of finite arbitrary length. We denote 0^n (i.e., n-bit string of all zeroes) by $\mathbf{0}$. For any element $X \in \{0,1\}^*$, $|X|$ denotes the number of bits in x. For any two elements $X, Y \in \{0,1\}^*$, $X \| Y$ denotes the concatenation of X followed by Y. For $X, Y \in \{0,1\}^n$, $X \oplus Y$ denotes the addition modulo 2 of X and Y. For any $X \in \{0,1\}^*$, parse X as $X = X_1 \| X_2 \| \ldots \| X_l$ where for each $i = 1, \ldots, l-1$, X_i is an element of $\{0,1\}^n$ and $1 \leq |X_l| \leq n$. We call each X_i a *block*. For a sequence of elements $(X_1, X_2, \ldots, X_s) \in \{0,1\}^*$, X_a^i denotes the a-th block of i-th element X_i.

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. For integers $1 \leq b \leq a$, $(a)_b$ denotes $a(a-1)\ldots(a-b+1)$, where $(a)_0 = 1$ by convention. $[q]$ refers to the set $\{1, \ldots, q\}$ and $[q_1, q_2]$ to the set $\{q_1, q_1+1 \ldots, q_2-1, q_2\}$.

2.1 Security Definitions

Pseudo Random Function (PRF) and Psuedo Random Permutation (PRP). A keyed function $\mathsf{F} : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$ with key space $\mathcal{K}$, domain $\mathcal{X}$ and range $\mathcal{Y}$ is a function for which $\mathsf{F}(K, X)$ shall be denoted by $\mathsf{F}_K(X)$. Given an oracle algorithm A that has oracle access to a function from $\mathcal{X}$ to $\mathcal{Y}$, makes at most q queries in time at most t, and returns a single bit, the prf-advantage of A against the family of keyed functions F is defined as

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

F is said to be a $(q, \ell, \sigma, t, \epsilon)$-secure PRF if $\mathbf{Adv}_{\mathsf{F}}^{\mathrm{PRF}}(q, \ell, \sigma, t) \triangleq \max_{\mathsf{A}} \mathbf{Adv}_{\mathsf{F}}^{\mathrm{PRF}}(\mathsf{A}) \leq \epsilon$, where the maximum is taken over all adversaries A that make q queries, with a maximum of ℓ data blocks in a single query and the total number of data blocks at most σ, with maximum running time t. Similarly, the prp-advantage of A against a family of keyed permutations E is defined as

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

E is said to be a (q, t, ϵ)-secure PRP if $\mathbf{Adv}_{\mathsf{E}}^{\mathrm{PRP}}(q, t) \triangleq \max_{\mathsf{A}} \mathbf{Adv}_{\mathsf{E}}^{\mathrm{PRP}}(\mathsf{A}) \leq \epsilon$, where maximum is taken over all adversaries A that make q queries and have running time at most t.

Message Authentication Code (MAC). Let $\mathcal{K}, \mathcal{N}, \mathcal{M}$ and $\mathcal{T}$ be four non-empty finite sets, $\mathsf{F} : \mathcal{K} \times \mathcal{N} \times \mathcal{M} \rightarrow \mathcal{T}$ be a nonce-based MAC. For $K \in \mathcal{K}$, let Auth_K be the authentication oracle, which takes as input $(N, M) \in \mathcal{N} \times \mathcal{M}$ and outputs $T = \mathsf{F}(K, N, M)$ and let Ver_K be the verification oracle, which takes as input $(N, M, T) \in \mathcal{N} \times \mathcal{M} \times \mathcal{T}$ and outputs 1 if $\mathsf{F}(K, N, M) = T$ and otherwise outputs 0. An authentication query (N, M) by an adversary A is called a **faulty query** if A has already queried to the first oracle with the same nonce but with a different message.

A (μ, q_m, q_v, t)-adversary against the unforgeability of F is an adversary A with oracle access to Auth_K and Ver_K such that it makes at most μ faulty authentication queries out of at most q_m authentication queries and q_v verification queries, with running time at most t. The adversary is said to be *nonce respecting* if $\mu = 0$ and nonce misusing if $\mu \geq 1$. However, the adversary may repeat nonces in its verification queries. A is said to *forge* F if for any of its verification queries (not obtained through a previous authentication query), the verification oracle returns 1. The advantage of A against the unforgeability of F is defined as

$$\mathbf{Adv}_{\mathsf{F}}^{\mathrm{MAC}}(\mathsf{A}) \triangleq \Pr\left[K \leftarrow_{\$} \mathcal{K} : \mathsf{A}^{\mathsf{Auth}_K(\cdot,\cdot), \mathsf{Ver}_K(\cdot,\cdot,\cdot)} \text{ forges }\right].$$

We write $\mathbf{Adv}_{\mathsf{F}}^{\mathrm{MAC}}(\mu, q_m, q_v, t) \triangleq \max_{\mathsf{A}} \mathbf{Adv}_{\mathsf{F}}^{\mathrm{MAC}}(\mathsf{A})$ where the maximum is taken over all (μ, q_m, q_v, t)-adversaries. In all of these definitions, we skip the parameter t, whenever we maximize over all unbounded adversaries.

Almost XOR Universal (AXU) Hash Function. Let $\mathcal{K}$ and $\mathcal{X}$ be two non-empty finite sets and H be a keyed function $\mathsf{H} : \mathcal{K}_h \times \mathcal{X} \rightarrow \{0,1\}^n$. Then, H is said to be an ϵ-almost xor universal hash function if for any distinct $X, X' \in \mathcal{X}$ and for any $Y \in \{0,1\}^n$,

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

We say that (X, X') is a colliding pair for a function H_{K_h} if $\mathsf{H}_{K_h}(X) = \mathsf{H}_{K_h}(X')$. H is said to be an ϵ-universal hash function if for any distinct $X, X' \in \mathcal{X}$,

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

Polyhash Function. A general algebraic hash function is a multivariate polynomial. Polyhash [29], one of the most popular examples of an algebraic hash function, is a univariate polynomial over the hash key K_h and its coefficients are the message blocks. For an n-bit hash key K_h, a message $M \in \{0,1\}^*$ is first padded with 10^* such that the number of bits in the padded message becomes a multiple of n. Let the padded message be $M^* = M_1 \| M_2 \| \ldots \| M_l$, where for each $i = 1, \ldots, l$, $|M_i| = n$. Then the PolyHash function is defined as follows:

$$\mathsf{PH}_{K_h}(M) = M_l K_h \oplus M_{l-1} K_h^2 \oplus \ldots \oplus M_1 K_h^l,$$

where l is the number of n-bit blocks of the padded message M^*. It is a well known result [17] that PolyHash is $\ell/2^n$-universal hash function, where ℓ is the maximum number of message blocks and the hash key is an element of the field $\mathrm{GF}(2^n)$.

2.2 A Brief Revisit to the expectation method

System and Distinguisher. Consider a computationally unbounded distinguisher A (hence assumed deterministic) that interacts with either of the possibly randomized stateful systems $\mathbf{S}_{\mathrm{re}}$ or $\mathbf{S}_{\mathrm{id}}$, after which it returns a single bit 0 or 1. For any such system $\mathbf{S}_{\mathrm{re}}$ or $\mathbf{S}_{\mathrm{id}}$, the interaction between A and the system defines an ordered sequence of queries and responses, $\tau = ((X_1, Y_1), (X_2, Y_2), \ldots, (X_q, Y_q))$ called a *transcript*, where X_i is the i-th query of A and Y_i is the corresponding response from the system. Let X_{re} (resp. X_{id}) be the random variable that takes a transcript resulting from the interaction between A and $\mathbf{S}_{\mathrm{re}}$ (resp. A and $\mathbf{S}_{\mathrm{id}}$). Then the advantage of A in distinguishing $\mathbf{S}_{\mathrm{re}}$ from $\mathbf{S}_{\mathrm{id}}$ is bounded from above by the statistical distance between the two random variables X_{re} and X_{id}, which is

$$\Delta(X_{\mathrm{re}}, X_{\mathrm{id}}) \overset{\Delta}{=} \sum_{\tau} \max\left\{0, \Pr\left[X_{\mathrm{id}} = \tau\right] - \Pr\left[X_{\mathrm{re}} = \tau\right]\right\}.$$

In the following, we briefly state the main result of the *Expectation Method* and show that the *coefficients-H technique* [32] is a special case of the expectation method. Both these techniques are used for bounding the information theoretic distinguishing advantage of two random systems as defined above.

expectation method. The expectation method was introduced by Hoang and Tessaro to derive a tight multi-user security bound of the key-alternating cipher [19]. Subsequently, this technique has been used for proving the multi-user security of the double encryption method in [20] and recently by Bose et al. to bound the multi-user security of $\mathsf{AES\text{-}GCM\text{-}SIV}$ [10]. This method is a generalization of coefficients-H technique. Let $\phi : \Theta \rightarrow [0, \infty)$ be a non-negative function which maps any attainable transcript to a non-negative real value. Suppose there is a set of good transcripts such that for any good transcript τ,

$$\frac{\Pr\left[X_{\mathrm{re}} = \tau\right]}{\Pr\left[X_{\mathrm{id}} = \tau\right]} \geq 1 - \phi(\tau). \tag{1}$$

The statistical distance between the two random variables X_{re} and X_{id} can then be bounded as

$$\Delta(X_{\mathrm{re}}, X_{\mathrm{id}}) \leq \mathbf{E}[\phi(X_{\mathrm{id}})] + \Pr[X_{\mathrm{id}} \in \Theta_{\mathrm{bad}}], \tag{2}$$

where Θ_{bad} is the set of all bad transcripts. In other words, the advantage of A in distinguishing $\mathbf{S}_{\mathrm{re}}$ from $\mathbf{S}_{\mathrm{id}}$ is bounded by $\mathbf{E}[\phi(X_{\mathrm{id}})] + \Pr[X_{\mathrm{id}} \in \Theta_{\mathrm{bad}}]$. coefficients-H technique can be seen as a simple corollary of the expectation method when ϕ is taken to be a constant function.

3 Design and Security Result of nEHtM and $\mathsf{CWC+}$

In this section we discuss the design and the security result of our proposed nonce-based message authentication code, called nEHtM and a nonce-based authenticated encryption scheme, called $\mathsf{CWC+}$. We begin our discussion with the EtM composition result that combines a standard encryption and a MAC scheme to achieve authenticated encryption.

3.1 Encrypt-then-MAC: Generic Composition Result

Bellare and Namprempre in [5] and Canetti and Krawczyk in [12] explored ways to combine standard encryption schemes with MACs to achieve authenticated encryption schemes. Their results yield three different types of combinations: (a) Encrypt-and-MAC ($\mathsf{E\&M}$), (b) MAC-then-Encrypt (MtE) and (c) Encrypt-then-MAC (EtM). In this paper we focus only on EtM.

Let $\mathcal{E} = (\mathcal{E}.\mathsf{KGen}, \mathcal{E}.\mathsf{Enc}, \mathcal{E}.\mathsf{Dec})$ be a nonce-based symmetric key encryption scheme and $\mathcal{I} = (\mathcal{I}.\mathsf{KGen}, \mathcal{I}.\mathsf{Tag}, \mathcal{I}.\mathsf{Ver})$ be a nonce-based message authentication code. The function $\mathcal{E}.\mathsf{Enc} : \mathcal{K}_e \times \mathcal{N} \times \mathcal{M} \to \mathcal{C}$ maps a tuple (K_e, N, M) to a ciphertext C and the decryption function $\mathcal{E}.\mathsf{Dec} : \mathcal{K}_e \times \mathcal{N} \times \mathcal{C} \to \mathcal{M} \cup \{\bot\}$ maps a legitimate tuple (K_e, N, C) to the corresponding message M and otherwise returns the error symbol $\bot$.

For the message authentication code $\mathcal{I}$, $\mathcal{I}.\mathsf{Tag} : \mathcal{K}_m \times \mathcal{N} \times \mathcal{D} \to \mathcal{T}$ maps a tuple (K_m, N, D) to a tag T and the verification function $\mathcal{I}.\mathsf{Ver} : \mathcal{K}_m \times \mathcal{N} \times \mathcal{M} \times \mathcal{T} \to \{\top, \bot\}$ maps a quadruple (K_e, N, D, T) to one of the two symbols $\{\top, \bot\}$ such that if T is the valid tag for the tuple (K_n, N, D) then the verification functions returns $\top$ (i.e., accept the message), otherwise it returns $\bot$ (i.e., reject the message).

Based on these two schemes, we define the EtM authenticated encryption scheme $\mathsf{AE}_{\mathcal{E},\mathcal{I}} = (\mathsf{AE.KGen}, \mathsf{AE.Enc}, \mathsf{AE.Dec})$ where the key-generation algorithm generates a random pair of keys $(K_e, K_m) \in \mathcal{K}_e \times \mathcal{K}_m$. The encryption and decryption algorithms are defined as follows:

$$\mathsf{AE.Enc}(K_e \| K_m, N, A, M) = \begin{cases} C \leftarrow \mathcal{E}.\mathsf{Enc}(K_e, N, M) \\ T \leftarrow \mathcal{I}.\mathsf{Tag}(K_m, N, A \| C) \end{cases}$$

$$\mathsf{AE.Dec}(K_e \| K_m, N, A, C, T) = \begin{cases} M \leftarrow \mathcal{E}.\mathsf{Dec}(K_e, N, C), \text{ if } Z = \top \\ \bot, \text{ if } Z = \bot \end{cases}$$

for $Z \leftarrow \mathcal{I}.\mathsf{Ver}(K_m, N, A \| C, T)$. We consider two security notions for the AE scheme: privacy and authenticity. The privacy advantage of AE is defined as follows:

$$\mathbf{Adv}^{\text{priv}}_{\mathsf{AE}}(\mathsf{A}) \triangleq \Pr[(K_e \times K_m) \leftarrow_{\$} (\mathcal{K}_e \times \mathcal{K}_m) : \mathsf{A}^{\mathsf{AE.Enc}_{(K_e,K_m)}} = 1] - \Pr[\mathsf{A}^{\$} = 1],$$

where the random oracle \$ takes (N, A, M) as input and returns $(C, T) \leftarrow_{\$} \{0,1\}^{|M|+\rho}$. We assume that the adversary A is nonce respecting, that is it does not make two queries with the same nonce.

If an adversary A interacts with the encryption and the decryption oracles of the AE, then the authenticity advantage of the AE is defined as follows:

$$\mathbf{Adv}^{\text{auth}}_{\mathsf{AE}}(\mathsf{A}) \triangleq \Pr[(K_e \times K_m) \leftarrow_{\$} (\mathcal{K}_e \times \mathcal{K}_m) : \mathsf{A}^{\mathsf{AE.Enc}_{(K_e,K_m)}, \mathsf{AE.Dec}_{(K_e,K_m)}} \text{ forges}],$$

where we say that the adversary A forges if the $\mathsf{AE.Dec}$ oracle returns a bit string (which is not $\bot$) for a query (N, A, C, T) such that (C, T) was not returned by the $\mathsf{AE.Enc}$ oracle as a result of the encryption query (N, A, M). Moreover, we assume that A can repeat nonces in decryption queries and can also use the nonces used in encryption queries.

The security of an AE scheme refers to the sum of its privacy and authenticity advantages. The privacy advantage of a nonce-based encryption scheme $\mathcal{E}$ that forms an AE with a MAC $\mathcal{I}$ is bound by the PRF advantage $\mathcal{E}$ and $\mathcal{I}$, while its authenticity advantage is bound by the forging advantage of the underlying $\mathcal{I}$. The achievement of a beyond birthday bound secure nonce-based AE scheme following the EtM paradigm thus requires a nonce respecting BBB secure nonce-based encryption scheme and a MAC mode that gives beyond birthday bound security for PRF-distinguishability and unforgeability (possibly in the nonce misuse model).

3.2 Encryption Modes used in Encrypt-then-MAC-based AE

A symmetric encryption scheme is generally defined through a pseudorandom number generator (PRNG) that takes a short master key K and an initial value or nonce N that generates a key stream $(S_1, S_2, \ldots)$. Then the ciphertext is generated from the plaintext and the key stream by applying the one time padding technique.

The counter mode of encryption (CTR) is a popular symmetric key encryption scheme, which gives birthday bound security in terms of the number of blocks, and is used as the underlying encryption scheme in AE constructions such as CWC [23], GCM [25], $\mathsf{GCM/2+}$ [3], $\mathsf{GCM\text{-}RUP}$ [4]. On the other hand Multi-EDM [40] and Multi-EDMD [40], which give an almost n-bit security, are used as the underlying encryption scheme in $\mathsf{OGCM1}$ [40] and $\mathsf{OGMC2}$ [40] respectively.

<u>Cipher-Based Encryption.</u> Cipher-based encryption [21] (CENC) is parametrized by a fixed non-negative integer w and so can be denoted as CENC_w. The PRNG of CENC_w takes a key K, a nonce ctr and a length l as input and outputs a sequence of fixed length key stream blocks, where the i-th key stream block is defined as

$$S_i \overset{\Delta}{=} \mathsf{E}_K(\mathsf{ctr} + j(w+1)) \oplus \mathsf{E}_K(\mathsf{ctr} + j(w+1) + i),\ j \in [0, l'-1], i \in [1, w],$$

where $l' = l/w$. The optimal security of CENC_w has been shown in [8] and it is used as the underlying encryption scheme of CHM and CIP AE constructions. An optimally secure nonce-based encryption mode $\mathsf{CENC}_{\max}$ [8], in which w is set to the maximum number of message blocks, is applied as the underlying encryption scheme of mGCM [8].

3.3 MACs used in Encrypt-then-MAC-based AEs

<u>Wegman-Carter MAC.</u> The Wegman-Carter (WC) MAC [39] is an early and popular nonce-based MAC that authenticates a message by masking its

hash value with a random number generated through a pseudorandom function applied on a nonce i.e.

$$\mathsf{WC}[\mathsf{F},\mathsf{H}](N,M) \overset{\Delta}{=} \mathsf{F}_K(N) \oplus \mathsf{H}_{k_h}(M).$$

The WC MAC provides $O(\epsilon q_v)$ security when nonces are never reused, where ϵ is the hash differential probability and q_v is the number of verification attempts. However, the construction has no security when the nonce repeats even once. For some constructions, the hash key is revealed and for others, a simple forgery is possible. Different instantiations of the pseudorandom function and hash function gives different instances of the WC MAC. The Wegman-Carter-Shoup (WCS) MAC [38] is a popular instantiation of WC MAC, where the pseudorandom function is replaced by a block cipher. WCS has been used as the underlying MAC in GCM, CHM and CIP. EDM and EDMD are used as instantiations of the PRF in WC MAC and the resultant MACs are used as the underlying MAC algorithms in $\mathsf{OGCM1}$ and $\mathsf{OGCM2}$ respectively. CWC MAC [23] (used as the MAC function in the CWC AE construction) is an another variant of the WC MAC where the pseduorandom function is replaced by a block cipher and the hash function is defined as $E_{K_2}(\mathsf{H}_{K_h}(M))$.

<u>Encrypted Wegman-Carter-Shoup.</u> The Encrypted Wegman-Carter-Shoup (EWCS) MAC [13] has been proposed as a remedy to the problem of nonce misuse security over the WC MAC. The EWCS MAC encrypts the output of the WCS MAC to generate the tag, and it is then used as the underlying MAC of $\mathsf{GCM/2+}$ construction. EWC gives a security of around $2^{n/2}$ when nonces do not repeat. An attacker can make approximately $2^{n/2}$ queries with distinct nonces but the same message and observe no collisions in the tag.

<u>Xor-Encrypt-Xor.</u> Xor-Encrypt-Xor (XEX) was originally proposed as a mode of designing a tweakable block cipher [37]. Luykx et al. [4] used it as the underlying MAC in $\mathsf{GCM\text{-}RUP}$. For a nonce N and a message M, XEX works as follows

$$\mathsf{XEX}[\mathsf{E},\mathsf{H}](N,M) \overset{\Delta}{=} \mathsf{E}_K(N \oplus \mathsf{H}_{K_h}(M)) \oplus \mathsf{H}_{K_h}(M).$$

XEX is secure upto the birthday bound when nonces do no repeat. It can be easily seen that a collision amongst the values of $N \oplus \mathsf{H}_{K_h}(M)$ leads to a forgery which can be easily detected by finding collision in the values $N \oplus T$.

EWCDM [13] and a single-keyed hash variant of $\mathsf{CLRW2}$ [24] are some possible alternatives of nonce-based MACs that can potentially be applied as the MAC function of any EtM-based AE mode. EWCDM has been proven to be secure upto approximately $2^{2n/3}$ queries when nonces do not repeat [13], and the single-keyed hash variant of $\mathsf{CLRW2}$ can be shown to be birthday bound secure in the nonce respecting setting.

It is to be noted that all these constructions has birthday bound PRF security as an attacker can make $2^{n/2}$ queries with distinct nonces but same message and observes no collision in the tag.

3.4 Security Result of nEHtM: A Nonce-Based Version of EHtM

The previous section demonstrates that the MACs used in the existing AE modes are not secure beyond the birthday bound when nonces repeat just once, making them unsuitable for use in designing an AE that is resilient in the faulty nonce model. This section introduces the *nonce-based Enhanced Hash-then-Mask* nEHtM and gives upto $2n/3$-bit unforgeability in faulty nonce model. The Enhanced Hash-then-Mask (EHtM) proposed by Minematsu [28], is the first BBB secure PRF-based probabilistic MAC that uses only an n-bit random salt and an n-bit PRF. nEHtM is structurally similar to EHtM, except that the random salt is replaced by a nonce and the PRF by a block cipher. Moreover, for the purpose of domain separation, we consider an $(n-1)$-bit nonce and an $(n-1)$-bit keyed hash function. For any message M and nonce N, nEHtM is defined as follows

$$\mathsf{nEHtM}[\mathsf{E}, \mathsf{H}_{K_h}](N, M) \overset{\Delta}{=} E_K(0\|N) \oplus \mathsf{E}_K(1\|(N \oplus \mathsf{H}_{K_h}(M))).$$

We now state Theorem 1, which bounds the unforgeability of nEHtM in the faulty nonce model. We also demonstrate a birthday bound forging attack on nEHtM when the number of faulty nonces reaches an order of $2^{n/2}$. The underlying idea of the attack is to form an alternating cycle of length 4 in the input of the block cipher; details may be found in Supplementary Sect. A.

Theorem 1. *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 $\mathsf{H} : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^{n-1}$ be an ϵ-AXU $(n-1)$-bit ϵ-AXU hash function. Let μ be a fixed parameter. Then the forging advantage for any (μ, q_m, q_v, t)-adversary against $\mathsf{nEHtM}[\mathsf{E}, \mathsf{H}]$ that makes authentication queries with at most μ faulty nonces is given by*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{nEHtM}[\mathsf{E},\mathsf{H}]}(\mu, q_m, q_v, t) \leq \mathbf{Adv}^{\mathrm{PRP}}_{\mathsf{E}}(\mu, q_m + q_v, t') + \frac{48q_m^3}{2^{2n}} + \frac{12q_m^4\epsilon}{2^{2n}} + \frac{12\mu^2 q_m^2}{2^{2n}} + \frac{q_m + 2q_v}{2^n} + \frac{4q_m^3\epsilon}{2^n} + (2q_m + q_v)\mu\epsilon + q_v\epsilon,$$

where the time parameter t' is of the order of $t + (q_m + q_v)t_{\mathsf{H}}$ and t_{H} is the time required for computing the hash function. Assuming $\epsilon \approx 2^{-(n-1)}$ and $q_m \leq \epsilon^{-1}$ simplifies this bound to

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{nEHtM}[\mathsf{Perm},\mathsf{H}]}(\mu, q_m, q_v, t) \leq \frac{72q_m^3}{2^{2n}} + \left(\frac{12\mu^2 q_m^2}{2^{2n}} + \frac{(4q_m + 2q_v)\mu}{2^n}\right) + \left(\frac{q_m + 4q_v}{2^n}\right).$$

The proof of this theorem is deferred until Sect. 6. The forging advantage of nEHtM for $\mu \leq 2^{n/3}$ and $q_m \leq 2^{2n/3}$ is thus given by

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{nEHtM}[\mathsf{Perm},\mathsf{H}]}(q_m, q_v, t) \leq \frac{13q_m}{2^{2n/3}} + \frac{4q_v}{2^{2n/3}}.$$

Remark 1. EHtM offers $3n/4$-bit security [18], whereas its nonce-based variant offers $2n/3$-bit security. This is because of the need to bound the number of multicollisions in the underlying hash function, for which the only source of randomness present in nEHtM is the hash key whereas EHtM also involves the random salts as an additional source of entropy.

3.5 CWC+: A beyond birthday bound variant of CWC

We have already seen that $\mathsf{CENC}_{\max}$ is a highly efficient optimally secure nonce respecting encryption scheme and nEHtM is a nonce-based MAC that is secure beyond the birthday bound in the faulty nonce model. Glueing them together using the EtM paradigm, we realize an authenticated encryption scheme, called CWC+, which gives a beyond the birthday bound security in the faulty nonce model. The encryption and decryption functions of CWC+ are shown in Fig. 3.1. The privacy and the authenticity advantages of CWC+ are stated in the following theorem, the proof of which is deferred until Sect. 7.

Theorem 2 (Privacy and Authenticity Bound of CWC+). *Let* $\mathsf{E} : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ *be a block cipher and* $\mathsf{Poly} : \{0,1\}^n \times \{0,1\}^* \to \{0,1\}^{n-1}$ *be the* $(n-1)$*-bit truncated PolyHash function which truncates the first bit of the PolyHash output. Let* ρ *and* μ *be two fixed parameters. Then the privacy advantage for any* $(q_e, q_d, \ell, \sigma, t)$*-nonce respecting adversary against* $\mathsf{CWC+}[\mathsf{E}, \rho]$ *is given by*

$$\mathbf{Adv}^{\mathrm{priv}}_{\mathsf{CWC+}[\mathsf{E},\rho]}(q_e, q_d, \ell, \sigma, t) \leq \mathbf{Adv}^{\mathrm{PRP}}_{\mathsf{E}}(\sigma + 2q, t') + \frac{105\sigma^3\ell}{2^{2n}} + \frac{6\sigma\ell}{2^n} + \frac{2q_d}{2^\rho} + \frac{2q_d\ell}{2^n}.$$

The authenticity advantage for any $(\mu, q_e, q_d, \ell, \sigma, t)$*-adversary against* $\mathsf{CWC+}[\mathsf{E}, \rho]$ *is given by*

$$\begin{aligned}\mathbf{Adv}^{\mathrm{auth}}_{\mathsf{CWC+}[\mathsf{E},\rho]}(\mu, q_e, q_d, \ell, \sigma, t) \leq \mathbf{Adv}^{\mathrm{PRP}}_{\mathsf{E}}(\sigma + 2q, t') &+ \frac{105\sigma^3\ell}{2^{2n}} + \frac{6\sigma\ell}{2^n} + \frac{2q_d}{2^\rho} + \frac{2q_d\ell}{2^n} \\ &+ \frac{(2q_e + q_d)2\ell\mu}{2^n} + \left(\frac{5\sigma\ell\mu}{2^n}\right)^2.\end{aligned}$$

We denote $q = q_e + q_d$*, the total number of encryption and decryption queries and* $t' = O(t + qt_{\mathsf{H}} + \sigma + 2q)$*, where* t_{H} *denotes the time for computing the hash function and* μ *denotes the total number faulty encryption queries.*

4 Mirror Theory

Mirror theory, introduced by Patarin in [33], is a technique to provide a lower bound for the number of solutions to a given system of linear (more precisely, affine) bivariate equations and non-equations in a finite field (e.g., $\mathrm{GF}(2^n)$). Solving a system of linear or affine equations is straightforward and a common problem in linear algebra. However, the problem starts complicating when non-equations are included. A special form of problems involving non-equations is to find distinct solutions to all the variables present in the system. If $Y_1, \ldots, Y_s$ are the variables, the system of non-equations $Y_i \oplus Y_j \neq \mathbf{0}$ for all $i \neq j$ essentially restricts the solutions to those in which all variables take distinct values. We call such a solution an *injective solution.* However, Patarin did not consider any other forms of non-equations [33–35]. This has been considered and termed as

```
**Algorithm** $\mathsf{CWC+.Enc}_K(N, A, M)$
    $L \leftarrow \mathsf{E}_K(\mathbf{0})$; $N' \leftarrow N \| 0^{n/4-1}$;
    $l \leftarrow \lceil |M|/n \rceil$;
    $S \leftarrow \mathsf{CENC}_{\max}(K, 0\|N', l)$;
    $C \leftarrow M \oplus \mathsf{first}(S, |M|)$;
    $\tilde{T} \leftarrow \mathsf{nEHtM}[\mathsf{E}, \mathsf{Poly}_{\mathsf{E}_K(\mathbf{0})}](N', C\|A)$;
    $T \leftarrow \mathsf{chop}_\rho(\tilde{T})$;
    return $(C, T)$
```

```
**Algorithm** $\mathsf{CWC+.Dec}_K(N, A, C, T)$
    $L = \mathsf{E}_K(\mathbf{0})$; $N' \leftarrow N \| 0^{n/4-1}$;
    $l \leftarrow \lceil |C|/n \rceil$;
    $\tilde{T}' \leftarrow \mathsf{nEHtM}[\mathsf{E}, \mathsf{Poly}_{\mathsf{E}_K(\mathbf{0})}](N', C\|A)$;
    if $\mathsf{chop}_\rho(\tilde{T}') \neq T$ then return $\perp$;
    $S \leftarrow \mathsf{CENC}_{\max}(K, N', l)$;
    $M \leftarrow C \oplus \mathsf{first}(S, |C|)$;
    return $M$
```

Fig. 3.1. Encryption and Decryption functions of $\mathsf{CWC+}$. $\mathsf{Poly}_{\mathsf{E}_K(\mathbf{0})}$ denotes the Poly-hash function with its n-bit hash key set to the encrypted value of $\mathbf{0}$. $\mathsf{first}(S, |M|)$ denotes the first $|M|$ bits in the sequence S. chop_ρ is a function that truncates the last $n - \rho$ bits of its input.

extended mirror theory in a recent work of Datta et al. [16]. In [16], the authors provided a lower bound on the number of injective solutions when the maximum component size $w_{\max}$ (a parameter that shall be defined soon) is three or less. This paper extends their analysis for an arbitrary $w_{\max}$.

Injective Solution of Equations. Let $G = (\mathcal{V} \triangleq \{Y_1, \ldots, Y_\alpha\}, \mathcal{S})$ be a simple acyclic graph with an edge-labelling function $\mathcal{L} : \mathcal{S} \to \{0,1\}^n$. For an edge $\{Y_i, Y_j\} \in \mathcal{S}$, we write $\mathcal{L}(\{Y_i, Y_j\}) = \lambda_{ij}$ (and so $\lambda_{ij} = \lambda_{ji}$). The system of equations induced by G, denoted $\mathcal{E}_G$, is then defined as:

$$\mathcal{E}_G \triangleq \{Y_i \oplus Y_j = \lambda_{ij};\ \{Y_i, Y_j\} \in \mathcal{S}\}. \tag{3}$$

That is, each vertex of G denotes a variable in the system of equations and each edge of G denotes an equation in $\mathcal{E}_G$. We denote the set of components in G by $\mathsf{comp}(G) = (\mathcal{C}_1, \ldots, \mathcal{C}_k)$, where k is the number of components in G. w_i denotes the size of (i.e. the number of vertices in) the component $\mathcal{C}_i$, $w_{\max}$ denotes the quantity $\max\{w_1, \ldots, w_k\}$ (also commonly denoted as ξ in Patarin's papers) and σ_i the sum $(w_1 + \cdots + w_i)$ with the convention that $\sigma_0 = 0$.

Definition 1. *With respect to the system of equations $\mathcal{E}_G$ (as defined above), an injective function $\Phi : \mathcal{V} \to \{0,1\}^n$ is said to be an* injective solution *if $\Phi(Y_i) + \Phi(Y_j) = \lambda_{ij}$ for all $\{Y_i, Y_j\} \in \mathcal{S}$.*

As the graph G is acyclic, there exists a unique path in the graph between any two vertices Y_s and Y_t in the same connected component, which shall be denoted by P_{st}. Adding all equations induced by the edges of any such path P_{st} gives

$$\mathcal{L}(P_{st}) := \sum_{e \in P_{st}} \mathcal{L}(e) = Y_s \oplus Y_t.$$

So, for an injective solution to exist, the graph G (along with the label function) must satisfy the following property:

NPL (non-zero path label): *For all paths P in graph G, $\mathcal{L}(P) \neq \mathbf{0}$.*

It may be noted here that the NPL condition formalizes the notion of non-degeneracy as mentioned in [33, 27]. The restriction on the graph to be acyclic implies that the equations are linearly independent (since otherwise, there is a possibility that the system becomes inconsistent).

Having identified the necessary condition for the existence of an injective solution to $\mathcal{E}_G$ corresponding to any simple edge-labeled undirected acyclic graph G, we now state the following claim due to Patarin [33], which gives a lower bound on the number of injective solutions to $\mathcal{E}_G$. Suppose G has α vertices and q edges. Patarin claimed that the number of injective solutions to $\mathcal{E}_G$ is at least $\frac{(2^n)_\alpha}{2^{nk}}$, provided $\sigma_k(w_{\max} - 1) \leq 2^n/64$. Unfortunately, the proof of this claim is unverifiable. [16] gives a detailed proof for the following lower bound on the number of injective solutions: $\frac{(2^n)_\alpha}{2^{nk}} \cdot (1 - \epsilon)$, with $\epsilon \approx 0$ and $\sigma_k^3 w_{\max}^2 \ll 2^{2n}$.

Injective Solution to a System of Equations and Non-Equations. An extended system involving a system of non-equations along with a system of equations shall now be examined. Let $G = (\mathcal{V} \triangleq \{Y_1, \ldots, Y_\alpha\}, \mathcal{S} \sqcup \mathcal{S}', \mathcal{L})$ be a simple undirected edge-labelled graph ($\mathcal{L}$ is a label function), whose edge set is partitioned into two disjoint sets $\mathcal{S}$ and $\mathcal{S}'$. As before, we simply write $\mathcal{L}(\{Y_i, Y_j\}) = \lambda_{ij}$ for all $\{Y_i, Y_j\} \in \mathcal{S}$ and $\mathcal{L}(\{Y_i, Y_j\}) = \lambda'_{ij}$ for all $\{Y_i, Y_j\} \in \mathcal{S}'$. Let such a graph G induce a system of equations and non-equations $\mathcal{E}_G$ as follows:

$$Y_i \oplus Y_j = \lambda_{ij} \ \forall \ \{Y_i, Y_j\} \in \mathcal{S}, \tag{4}$$
$$Y_i \oplus Y_j \neq \lambda'_{ij} \ \forall \ \{Y_i, Y_j\} \in \mathcal{S}', \tag{5}$$

For a system of equations and non-equations $\mathcal{E}_G$, an injective function $\Phi : \mathcal{V} \to \{0,1\}^n$ is said to be an *injective solution function* if $\Phi(Y_i) \oplus \Phi(Y_j) = \lambda_{ij}$ for all $\{Y_i, Y_j\} \in \mathcal{S}$ and $\Phi(Y_i) \oplus \Phi(Y_j) \neq \lambda'_{ij}$ for all $\{Y_i, Y_j\} \in \mathcal{S}'$.

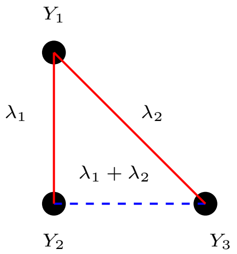

Fig. 4.1. $\mathcal{E}_G \triangleq \{Y_1 \oplus Y_2 = \lambda_1, Y_1 \oplus Y_3 = \lambda_2, Y_2 \oplus Y_3 \neq \lambda_1 \oplus \lambda_2\}$. The continuous red edges represent equations and the dashed blue edges represent non-equations. Clearly, the system of equations and non-equations is inconsistent.

Good Graphs. We shall first investigate the case when $\mathcal{E}_G$ has at least one solution. To ensure this, the subgraph $G^{=} \triangleq (\mathcal{V}, \mathcal{S}, \mathcal{L}_{|\mathcal{S}})$, where $\mathcal{L}_{|\mathcal{S}}$ is the function $\mathcal{L}$ restricted over the set $\mathcal{S}$, must

(i) be acyclic (i.e. **No Cycle** or **NC**)
(ii) satisfy the **NPL** condition and
(iii) satisfy the **NCL (non-zero cycle label)** property which says that *for all cycles C in G such that the edge set of C contains exactly one non-equation edge $e' \in \mathcal{S}'$, $\mathcal{L}(C) \neq \mathbf{0}$* (see Fig.4.1 for an example).

If a graph G satisfies the above three conditions (i)-(iii), it is said to be a **good graph**. In [16], authors have proved the following lower bound for $w_{\max} = 3$. Let $G = (\mathcal{V}, \mathcal{S} \sqcup \mathcal{S}', \mathcal{L})$ be a good graph with $|\mathcal{V}| = \alpha, |\mathcal{S}| = q_m, |\mathcal{S}'| = q_v$. Let $\mathsf{comp}(G^{=}) = (\mathcal{C}_1, \ldots, \mathcal{C}_k)$ with $|\mathcal{C}_i| = w_i$ (≤ 3) and $\sigma_i = (w_1 + \cdots + w_i)$. Let $\mathcal{Z} \subseteq \{0,1\}^n$ such that $|\{0,1\}^n \setminus \mathcal{Z}| = c$. The total number of injective solutions (each solution is chosen from the set $\mathcal{Z}$) for the induced system of equations and non-equations $\mathcal{E}_G$ is at least:

$$\frac{(2^n)_\alpha}{2^{nk}}\left(1 - \frac{5k^3}{2^{2n}} - \frac{q_v + c\alpha}{2^{n-1}}\right). \tag{6}$$

Observe that $q_v + c\alpha$ is the number of non-equations, considering univariate non-equations arising from the constraint of each solution being from the set of size $2^n - c$. Now we state our theorem, which generalizes this result for any $w_{\max}$.

Theorem 3. *Let $G = (\mathcal{V}, \mathcal{S} \sqcup \mathcal{S}', \mathcal{L})$ be a good graph with α vertices such that $|\mathcal{S}| = q_m, |\mathcal{S}'| = q_v$. Let $\mathsf{comp}(G^{=}) = (\mathcal{C}_1, \ldots, \mathcal{C}_k)$ and $|\mathcal{C}_i| = w_i$, $\sigma_i = (w_1 + \cdots + w_i)$. Then the total number of injective solutions chosen from a set $\mathcal{Z}$ of size $2^n - c$, for some $c \geq 0$, for the induced system of equations and non-equations $\mathcal{E}_G$ is at least:*

$$\frac{(2^n)_\alpha}{2^{nq}}\left(1 - \sum_{i=1}^{k} \frac{6\sigma_{i-1}^2 \binom{w_i}{2}}{2^{2n}} - \frac{2(q_v + c\alpha)}{2^n}\right), \tag{7}$$

provided $\sigma_k w_{\max} \leq 2^n/4$.

Proof. We give here a brief sketch of the proof. A detailed proof of the theorem can be found in Supplementary Sect. B. The proof proceeds by counting the number of solutions in each of the k components. We denote $\tilde{w}_{ij}$ to be the number of edges from $\mathcal{S}'$ connecting vertices between i-th and j-th component of $G^{=}$ and w'_i to be the number of edges in $\mathcal{S}'$ incident on $v_i \in \mathcal{V} \backslash G^{=}(\mathcal{V})$. It is easy to see that the number of solutions for the first component is exactly $(2^n - cw_1)$. We fix a solution and count the number of solutions for the second component which is $(2^n - w_1 w_2 - \tilde{w}_{1,2} - cw_2)$ as it must discard (i) w_1 values $(y_{i_1}, \ldots, y_{i_{w_1}})$ from the first component, (ii) $w_1(w_2 - 1)$ values $(y_{i_1} \oplus \mathcal{L}(P_j), \ldots, y_{i_{w_1}} \oplus \mathcal{L}(P_j))$ for all possible paths P_j from a fixed vertex to any other vertex in the second component and (iii) $cw_2 + \tilde{w}_{12}$ values to compensate for the fact that the set of values is no longer a group. In general, the total number of solutions for the i-th component is at least $\prod\limits_{i=1}^{k}\left(2^n - \sigma_{i-1} w_i - \sum\limits_{j=1}^{i-1} \tilde{w}_{ij} - cw_i\right)$. Suppose there are

k' vertices that do not belong to the set of vertices of the subgraph $G^{=}$. Fix such a vertex Y_{σ_k+i} and let us assume that w'_{σ_k+i} blue dashed edges are incident on it. If y_{σ_k+i} is a valid solution to the variable Y_{σ_k+i}, then we must have (a) y_{σ_k+i} should be distinct from the previous σ_k assigned values, (b) y_{σ_k+i} should be distinct from the $(i-1)$ values assigned to the variables that do not belong to the set of vertices of the subgraph $G^{=}(\mathcal{V})$ and (c) y_{σ_k+i} should not take those w'_{σ_k+i} values.
Therefore, the total number of solutions is at least

$$h_\alpha \geq \prod_{i=1}^{k} \left(2^n - \sigma_{i-1}w_i - \sum_{j=1}^{i-1} \tilde{w}_{ij} - cw_i\right) \cdot \prod_{i \in [k']} (2^n - \sigma_k - i + 1 - w'_{\sigma_k+i}). \quad (8)$$

Let us denote $(\tilde{w}_{i1} + \ldots + \tilde{w}_{i,i-1})$ by p_i and $(w'_{\sigma_k+1} + \ldots + w'_{\sigma_k+k'})$ by q''_v. After a simple algebraic calculation on Eqn. (8), we obtain

$$h_\alpha \frac{2^{nq_m}}{(2^n)_\alpha} \geq \underbrace{\prod_{i=1}^{k} \frac{(2^n - \sigma_{i-1}w_i - p_i - cw_i)2^{n(w_i-1)}}{(2^n - \sigma_{i-1})_{w_i}}}_{\mathsf{D.1}} \left(1 - \frac{2q''_v}{2^n}\right). \quad (9)$$

Let us denote the expression $\left(\binom{w_i}{2}\sigma_{i-1}^2 + \binom{w_i}{2}(w_i-1)\sigma_{i-1} + \binom{w_i}{2}\frac{(w_i-2)(3w_i-1)}{12}\right)$ by A_i. Expanding $(2^n - \sigma_{i-1})_{w_i}$ along with some simple computations on $\mathsf{D.1}$ gives

$$\begin{aligned} \mathsf{D.1} &\geq \prod_{i=1}^{k} \left(1 - \frac{A_i}{2^{2n} - 2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) + A_i} - \frac{2^n(p_i + cw_i)}{2^{2n} - 2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) + A_i}\right) \\ &\overset{(4)}{\geq} \prod_{i=1}^{k} \left(1 - \frac{2A_i}{2^{2n}} - \frac{2(p_i + cw_i)}{2^n}\right) \overset{(5)}{\geq} \left(1 - \sum_{i=1}^{k} \frac{6\sigma_{i-1}^2\binom{w_i}{2}}{2^{2n}} - \sum_{i=1}^{k} \frac{2(p_i + cw_i)}{2^n}\right) \\ &\overset{(6)}{\geq} \left(1 - \sum_{i=1}^{k} \frac{6\sigma_{i-1}^2\binom{w_i}{2}}{2^{2n}} - \frac{2q'_v}{2^n} - \frac{2c\alpha}{2^n}\right), \quad (10) \end{aligned}$$

where (4) follows from the fact that $2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) - A_i \leq 2^{2n}/2$, which holds true when $\sigma_k w_{\max} \leq 2^n/4$, (5) holds true due to the fact that $A_i \leq 3\sigma_{i-1}^2\binom{w_i}{2}$ and (6) holds true as we denote $(p_1 + \ldots + p_k) = q'_v$, the total number of blue dashed edges across the components of $G^{=}$ and $w_1 + \ldots + w_k \leq \alpha$. Finally, from Eqn. (9), Eqn. (10) and $q_v = q'_v + q''_v$, the result follows. □

5 Mutlicollision in Universal Hash Function

In this section, we study the muticollision advantage of a universal hash function. Suppose H_{K_h} is an ϵ universal hash function where the hash key K_h is chosen uniformly at random from the hash-key space. For any q distinct messages

$M_1, \ldots, M_q$, the probability that there exist $i \neq j$, such that M_i and M_j collide under the hash function H_{K_h} is at most $\epsilon\binom{q}{2}$ (by the union bound). Extending this result for multicollisions, we say that $(M_1, \ldots, M_\xi)$ is a ξ-*multicollision tuple* for H_{K_h} if $\mathsf{H}_{K_h}(M_1) = \mathsf{H}_{K_h}(M_2) = \cdots = \mathsf{H}_{K_h}(M_\xi)$. When H_{K_h} is a ξ-wise independent hash function [39] the probability that a ξ-tuple $(M_1, \ldots, M_\xi)$ is a ξ-multicollision tuple for H_{K_h} is $1/2^{n(\xi-1)}$. Clearly, this cannot be concluded for a universal hash function. In fact, one can easily construct a ξ-tuple of messages such that the multicollision probability under the PolyHash function is $\ell/2^n$.

In the following, we now provide a bound (better than $\epsilon\binom{q}{2}$) on the existence of a multicollision tuple for any given q messages.

Theorem 4 (Multicollision Theorem). *Let $X_1, \ldots, X_q$ be q distinct messages and H_{K_h} be an ϵ-universal hash function. Then for $\xi \in \mathbb{N}$, the probability that a $(\xi+1)$-multicollision tuple exists in this set of messages is no more than $q^2\epsilon/2\xi$.*

Proof. Let us denote the required probability by P and set $Z_i = \mathsf{H}_{K_h}(X_i)$, $i \in [q]$. Also let $\mathbf{X}$ denote a $(\xi+1)$-tuple $(X_1, \ldots, X_{\xi+1}) \in \mathcal{V}^{\xi+1}$. Consider the graph $G = (\mathcal{V}, \mathcal{S})$ whose vertex set $\mathcal{V}$ contains each of the q messages. An edge between two nodes exists in $\mathcal{S}$ if and only if the hash values of the corresponding messages collide. Therefore, the event $\mathsf{H}_{K_h}(X_1) = \ldots = \mathsf{H}_{K_h}(X_{\xi+1})$ boils down to the existence of a clique of size $\xi+1$ in G. Due to Lemma 1, if G has $q^2/2\xi$ edges, then any collection of $\xi+1$ vertices of the q vertices in $\mathcal{V}$ must contain at least one pair which is in $\mathcal{S}$. i.e. there must exist $\{v_1, \ldots, v_s\} \subseteq [q]$, for $s = q^2/\xi$, such that

$$Z_1 = Z_2 = \ldots = Z_{\xi+1} \Rightarrow Z_{v_1} = Z_{v_2} \vee Z_{v_3} = Z_{v_4} \vee \ldots \vee Z_{v_s-1} = Z_{v_s}, \tag{11}$$

Therefore, the probability P is:

$$\max_{\mathbf{X}} \Pr\left[K_h \leftarrow_{\$} \{0,1\}^n : \exists i_1, \cdots, i_\xi \in [q], \mathsf{H}_{K_h}(X_{i_1}) = \cdots = \mathsf{H}_{K_h}(X_{i_\xi})\right]$$

$$\leq \Pr[Z_{v_1} = Z_{v_2} \vee \ldots \vee Z_{v_s-1} = Z_{v_s}] \leq \sum_{i=1}^{s/2} \Pr[Z_{v_i} = Z_{v_{i+1}}] \leq \frac{s\epsilon}{2} = \frac{q^2\epsilon}{2\xi}.$$

Lemma 1. *Let $q, \xi \in \mathbb{N}$. Then for any set $\mathcal{V}$ with $|\mathcal{V}| = q$, there exists a graph $G = (\mathcal{V}, \mathcal{S})$ with $|\mathcal{S}| = \left\lceil \frac{q^2}{2\xi} \right\rceil$ such that any collection C of $\xi+1$ vertices has at least one edge in $\mathcal{S}$ joining two vertices in C.*

Proof. Divide the q vertices into ξ subcollections of size $\left\lceil \frac{q}{\xi} \right\rceil$ each, the last subcollection possibly containing a lesser number of vertices. Construct $\mathcal{S}$ by adding in it, all the edges required to form a clique $C_i, i \in [\xi]$ out of each of the ξ subcollections. Thus, there are at most $\xi \cdot \binom{\lceil \frac{q}{\xi} \rceil - 1}{2}$ edges in all the ξ cliques. Observe that,

$$\xi \cdot \binom{\left\lceil \frac{q}{\xi} \right\rceil - 1}{2} < \xi \cdot \binom{\frac{q}{\xi}}{2} \leq \frac{q^2}{2\xi} \leq \left\lceil \frac{q^2}{2\xi} \right\rceil.$$

Hence, $\mathcal{S}$ must contain more edges, distinct from those involved in the ξ cliques, which must exist between at least one pair of vertices in different cliques C_i and C_j $(i \neq j)$. Since there are $\xi + 1$ vertices in C and a total of ξ cliques C_i formed so far in G_2, it can thus be inferred from the pigeonhole principle that at least one clique C_i contains more than one edge from $\mathcal{S}$, making clear the existence of an edge from $\mathcal{S}$ in C. □

6 Proof of Theorem 1

In this section, we prove Theorem 1. We shall often refer to the construction $\mathsf{nEHtM}[\mathsf{E}, \mathsf{H}]$ as simply nEHtM when the underlying primitives are assumed to be understood.

The first step of the proof is the standard switch from the computational setting to the information theoretic one by replacing the block cipher E_K with an n-bit uniform random permutation Π at the cost of $\mathbf{Adv}_{\mathsf{E}}^{\mathrm{PRP}}(q_m + q_v, t')$, where $t' = O(t + (q_m + q_v)t_H)$ and t_H is the time required for computing the hash function. Let us denote this modified construction as $\mathsf{nEHtM}^*[\Pi, \mathsf{H}]$. Hence,

$$\mathbf{Adv}_{\mathsf{nEHtM}}^{\mathrm{MAC}}(q_m, q_v, t) \leq \mathbf{Adv}_{\mathsf{E}}^{\mathrm{PRP}}(q_m + q_v, t') + \underbrace{\mathbf{Adv}_{\mathsf{nEHtM}^*}^{\mathrm{MAC}}(q_m, q_v, t)}_{\delta^*}. \tag{12}$$

To get an upper bound for δ^*, we consider a perfect random oracle Rand, which on input (N, M) returns T, sampled uniformly at random from $\{0,1\}^n$, and an oracle Rej which always returns $\perp$ (i.e., rejects) for all inputs (N, M, T). Now, due to [13, 18, 16] we have

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

where the maximum is taken over all non-trivial distinguishers D. This formulation allows us to apply the expectation method [19, 10] to prove that

$$\delta^* \leq \frac{48q_m^3}{2^{2n}} + \frac{12q_m^4\epsilon}{2^{2n}} + \frac{12\mu^2 q_m^2}{2^{2n}} + \frac{q_m + 2q_v}{2^n} + \frac{4q_m^3\epsilon}{2^n} + (2q_m + q_v)\mu\epsilon + q_v\epsilon. \tag{13}$$

Attack Transcript. Henceforth, we fix a deterministic non-trivial (i.e., one that makes no repeated queries) distinguisher D that interacts with either (1) the real oracle $(\mathsf{TG}[\Pi, \mathsf{H}_{K_h}], \mathsf{VF}[\Pi, \mathsf{H}_{K_h}])$ for a uniform random permutation Π 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 (authentication) oracle with at most μ faulty nonces and at most q_v queries to its right (verification) oracle, and returning a single bit. Then

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

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

be the list of authentication queries made by D and the corresponding responses it receives. Also let

$$\tau_v \triangleq \{(N'_1, M'_1, T'_1, b'_1), (N'_2, M'_2, T'_2, b'_2), \ldots, (N'_{q_v}, M'_{q_v}, T'_{q_v}, b'_{q_v})\}$$

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

A transcript τ' is said to be an *attainable transcript* (with respect to D) if the probability of realizing this transcript in the ideal world is non-zero. It must be noted that since attainability is with respect to the ideal world, any verification query (N'_i, M'_i, T'_i, b'_i) even in an attainable transcript $\tau' = (\tau, K_h)$ is such that $b'_i = \bot$. We denote Θ to be the set of all attainable transcripts and X_{re} and X_{id} to be the random variables that take an extended transcript τ' induced by the real world and the ideal world respectively.

6.1 Definition and Probability of Bad Transcripts

In this section, we define and bound the probability of bad transcripts in the ideal world. For notational simplicity, we denote $N_i \oplus \mathsf{H}_{K_h}(M_i)$ as X_i. Note that X_i is an $n-1$ bit string.

Definition 2 (Bad Transcript). *Given a paramter $\xi \in \mathbb{N}$, where $\xi \geq \mu$, an attainable transcript $\tau' = (\tau_m, \tau_v, K_h)$ is called a* **bad** *transcript if any one of the following holds:*

- $\mathsf{B1}$ *: $\exists\ i \in [q_m]$ such that $T_i = \mathbf{0}$.*
- $\mathsf{B2}$ *: $\exists\ i \neq j, j \neq k$ such that $N_i = N_j$ and $X_j = X_k$.*
- $\mathsf{B3}$ *: $\{i_1, \ldots, i_{\xi+1}\} \subseteq [q_m]$ such that $X_{i_1} = X_{i_2} = \ldots = X_{i_{\xi+1}}$ (the optimal value of ξ shall be determined later in the proof).*
- $\mathsf{B4}$ *$\exists\ a \in [q_v]$, $\exists\ i \in [q_m]$ such that $N_i = N'_a$, $X_i = X'_a$ and $T_i = T'_a$.*

We denote by Θ_{bad} (resp. Θ_{good}) the set of bad (resp. good) transcripts. We bound the probability of bad transcripts in the ideal world as follows.

Lemma 2. *Let X_{id} and Θ_{bad} be defined as above. Then*

$$\Pr[X_{\mathrm{id}} \in \Theta_{\mathrm{bad}}] \leq \epsilon_{\mathrm{bad}} = \frac{q_m}{2^n} + \frac{q_m^2 \epsilon}{2\xi} + (2q_m + q_v)\mu\epsilon + q_v\epsilon.$$

Proof. By the union bound,

$$\Pr[X_{\mathrm{id}} \in \Theta_{\mathrm{bad}}] \leq \Pr[\mathsf{B1}] + \Pr[\mathsf{B2}] + \Pr[\mathsf{B3}] + \Pr[\mathsf{B4}]. \tag{14}$$

In the following, we bound the probabilities of all the bad events individually. The lemma will follow by adding the individual bounds. Clearly,

$$\Pr[\mathsf{B1}] \leq \frac{q_m}{2^n}. \tag{15}$$

Bounding B2. Let $\mathcal{F}$ be the set of all query indices i for which there is a $j \neq i$ such that $N_i = N_j$. It is easy to see that $|\mathcal{F}| \leq 2\mu$. Event $\mathsf{B2}$ occurs if for some $j \in \mathcal{F}$, $\mathsf{H}_{K_h}(M_j) = N_k \oplus \mathsf{H}_{K_h}(M_k)$ for some $k \neq j$. For any such fixed i, j, k, the probability of the event is at most ϵ. The number of such choices of (j, k) is at most $2\mu q_m$. Hence,

$$\Pr[\mathsf{B2}] \leq 2\mu q_m \epsilon. \tag{16}$$

Bounding B3. Event $\mathsf{B3}$ occurs if there exist $\xi + 1$ distinct authentication query indices $\{i_1, \ldots, i_{\xi+1}\} \subseteq [q_m]$ such that $X_{i_1} = \ldots = X_{i_{\xi+1}}$. This event is thus a $(\xi+1)$-multicollision on the ϵ universal hash function mapping (N, M) to $\mathsf{H}_{K_h}(M) \oplus N$ (as H_{K_h} is an ϵ-almost-xor universal). Therefore, by Theorem 4:

$$\Pr[\mathsf{B3}] \leq q_m^2 \epsilon / 2\xi. \tag{17}$$

Bounding B4. For some $a \in [q_v]$ and $i \in [q_m]$, if $N_i = N'_a$, $X_i = X'_a$ and $T_i = T'_a$, then $M_i \neq M'_a$ (as the adversary does not make any trivial query). Hence the probability that $X_i = X'_a$ holds is at most ϵ. Now, for any a, there can be at most $(\mu + 1)$ indices i such that $N_i = N'_a$. Hence, the required probability is bounded as

$$\Pr[\mathsf{B4}] \leq (\mu + 1) q_v \epsilon. \tag{18}$$

The proof follows from Eqn. (14)-Eqn. (18). □

6.2 Analysis of Good Transcripts

In this section, we show that for a good transcript $\tau' = (\tau, K_h)$, realizing τ' is almost as likely in the real world as in the ideal world.

Consider a good transcript $\tau' = (\tau_m, \tau_v, K_h)$. Since in the ideal world the authentication oracle is perfectly random and the verification oracle always rejects,

$$\Pr[X_{\mathrm{id}} = \tau'] = \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{2^{n q_m}} \tag{19}$$

We must now lower bound $\Pr[X_{\mathrm{re}} = \tau']$ i.e., the probability of getting τ' in the real world. We say that a permutation Π is *compatible with* τ_m (respectively with τ_v) if (A) (respectively (B)) holds.

$$\text{(A) } \forall i \in [q_m], \Pi(\widehat{N}_i) \oplus \Pi(\widehat{X}_i) = T_i, \text{ (B) } \forall a \in [q_v], \Pi(\widehat{N'}_a) \oplus \Pi(\widehat{X'}_a) \neq T'_a,$$

where $\widehat{N}_i = 0\|N_i$, $\widehat{X}_i = 1\|X_i$, $\widehat{N'}_a = 0\|N'_a$ and $\widehat{X'}_a = 1\|X'_a$. We simply say that $\mathsf{\Pi}$ is compatible with $\tau = (\tau_m, \tau_v)$ if it is compatible with τ_m and τ_v. We denote by $\mathsf{Comp}(\tau)$ the set of permutations $\mathsf{\Pi}$ that are compatible with τ. Therefore,

$$\begin{aligned}\Pr[X_{\mathrm{id}} = \tau'] &= \frac{1}{|\mathcal{K}_h|} \cdot \Pr[\mathsf{\Pi} \leftarrow_{\$} \mathsf{Perm} : \mathsf{\Pi} \in \mathsf{Comp}(\tau)] \\ &= \frac{1}{|\mathcal{K}_h|} \cdot \underbrace{\Pr[\mathsf{\Pi}(\widehat{N}_i) \oplus \mathsf{\Pi}(\widehat{X}_i) = T_i, \mathsf{\Pi}(\widehat{N'}_a) \oplus \mathsf{\Pi}(\widehat{X'}_a) \neq T'_a]}_{\mathsf{P}_{mv}}. \quad (20)\end{aligned}$$

We refer to the system of equations as “*authentication equations*” as they involve only the authentication queries and to the system of non-equations as “*verification non-equations*” as they involve only the verification queries. We denote the system of authentication equations by $\mathcal{E}_m$ and the system of verification non-equations by $\mathcal{E}_v$.

$$(\mathcal{E}_m) = \begin{cases} \mathsf{\Pi}(\widehat{N}_1) \oplus \mathsf{\Pi}(\widehat{X}_1) = T_1 \\ \mathsf{\Pi}(\widehat{N}_2) \oplus \mathsf{\Pi}(\widehat{X}_2) = T_2 \\ \vdots \\ \mathsf{\Pi}(\widehat{N}_{q_m}) \oplus \mathsf{\Pi}(\widehat{X}_{q_m}) = T_{q_m} \end{cases} \qquad (\mathcal{E}_v) = \begin{cases} \mathsf{\Pi}(\widehat{N'}_1) \oplus \mathsf{\Pi}(\widehat{X'}_1) \neq T'_1 \\ \mathsf{\Pi}(\widehat{N'}_2) \oplus \mathsf{\Pi}(\widehat{X'}_2) \neq T'_2 \\ \vdots \\ \mathsf{\Pi}(\widehat{N'}_{q_v}) \oplus \mathsf{\Pi}(\widehat{X'}_{q_v}) \neq T'_{q_v} \end{cases}$$

Equation and Non-Equation Inducing Graph. From the above system of bivariate affine equations and non-equations, we induce the edge-labelled undirected graph $G_{\tau'} = (\mathcal{V}, \mathcal{S} \sqcup \mathcal{S}')$, where the set of nodes $\mathcal{V}$ is the set of variables $\{Y_1, \ldots, Y_\alpha\}$, $\mathcal{S}$ is the set of edges corresponding to each authentication equation and $\mathcal{S}'$ is the set of edges corresponding to each verification non-equation. Moreover, if there is an authentication equation $Y_s \oplus Y_t = T_i$, then the corresponding edge $\{Y_s, Y_t\} \in \mathcal{S}$ is labelled T_i. Similarly, if there is a verification non-equation $Y_s \oplus Y_t \neq T'_i$, then the corresponding edge $\{Y_s, Y_t\} \in \mathcal{S}'$ is labelled T'_i. Moreover, $G^{=}_{\tau'} = (\mathcal{V}, \mathcal{S})$ is the subgraph of $G_{\tau'}$.

The proof of the following claim can be found in Supplementary Sect. C.

Claim 1. *If the transcript τ' is good, then the induced graph $G_{\tau'}$ is valid.*

Suppose there are k components in the subgraph $G^{=}_{\tau'}$ and the size of the i-th component is W_i. Thus, W_i is a random variable, and so is $W_{\max}$, which denotes the size of the largest component. It is easy to see that $W_{\max} \leq \xi$. As the graph $G_{\tau'}$ is valid (follows from Claim 1), we assume $\xi \leq 2^n/8q_m$ (from the condition of Theorem 3), which allows us to apply Theorem 3 with $c = 0$ to obtain,

$$\mathsf{P}_{mv} \geq \frac{1}{2^{nq_m}} \cdot \left(1 - \sum_{i=1}^{k} \frac{6\sigma_{i-1}^2 \binom{W_i}{2}}{2^{2n}} - \frac{2q_v}{2^n}\right). \quad (21)$$

Therefore, Eqn. (19)-Eqn. (21) imply that the ratio $\frac{\Pr[X_{\mathrm{re}}=\tau']}{\Pr[X_{\mathrm{id}}=\tau']}$ is no less than

$$\left(1-\sum_{i=1}^{k}\frac{6\sigma_{i-1}^2\binom{W_i}{2}}{2^{2n}}-\frac{2q_v}{2^n}\right)\overset{(1)}{\geq} 1-\underbrace{\left(\sum_{i=1}^{k}\frac{24q_m^2\binom{W_i}{2}}{2^{2n}}+\frac{2q_v}{2^n}\right)}_{\phi(\tau')}, \tag{22}$$

where (1) follows due to the inequality $\sigma_{i-1} \leq 2q_m$.

We now compute the expectation of $\phi(X_{\mathrm{id}})$ as follows.

$$\mathbf{E}\left[\left(\sum_{i=1}^{k}\frac{24q_m^2\binom{W_i}{2}}{2^{2n}}+\frac{2q_v}{2^n}\right)\right]=\left(\frac{2q_v}{2^n}+\frac{24q_m^2}{2^{2n}}\mathbf{E}\left[\sum_{i=1}^{k}\binom{W_i}{2}\right]\right). \tag{23}$$

Let $\tilde{W}_i = W_i - 1$ and therefore,

$$\mathbf{E}\left[\sum_{i=1}^{k}\binom{W_i}{2}\right]=\mathbf{E}\left[\sum_{i=1}^{k}\binom{\tilde{W}_i}{2}\right]+\mathbf{E}\left[\sum_{i=1}^{k}\tilde{W}_i\right]\overset{(2)}{\leq}\mathbf{E}\left[\sum_{i=1}^{k}\binom{\tilde{W}_i}{2}\right]+2q_m. \tag{24}$$

where (2) holds as $(\tilde{W}_1+\ldots\tilde{W}_k)=\sigma_k-k\leq 2q_m$. Let us consider the following two indicator random variables

$$I_{ij}=\begin{cases}1, & \text{if } X_i=X_j\\ 0, & \text{otherwise}\end{cases}\qquad \tilde{I}_{ij}=\begin{cases}1, & \text{if } N_i=N_j\\ 0, & \text{otherwise.}\end{cases}$$

Therefore,

$$\begin{aligned}\mathbf{E}\left[\sum_{i=1}^{k}\binom{\tilde{W}_i}{2}\right] &\overset{(3)}{=}\sum_{i\neq j}^{q_m}\mathbf{E}[I_{ij}]+\sum_{i\neq j}^{\mu}\mathbf{E}[\tilde{I}_{ij}]\\ &\overset{(4)}{=}\sum_{i\neq j}^{q_m}\Pr[\mathsf{H}_{K_h}(M_i)\oplus\mathsf{H}_{K_h}(M_j)=N_i\oplus N_j]+\mu^2/2\\ &\overset{(5)}{\leq}\binom{q_m}{2}\epsilon+\mu^2/2\leq q_m^2\epsilon/2+\mu^2/2,\end{aligned} \tag{25}$$

where (3) holds due to the linearity of expectation, (4) holds from the definition of the indicator random variable and (5) holds from the ϵ-almost-xor universal probability of the underlying hash function. Therefore, from Eqn. (23)-Eqn. (25), we have

$$\mathbf{E}[\phi(X_{\mathrm{id}})]\leq\left(\frac{12q_m^4\epsilon}{2^{2n}}+\frac{12\mu^2q_m^2}{2^{2n}}+\frac{48q_m^3}{2^{2n}}+\frac{2q_v}{2^n}\right). \tag{26}$$

Finalizing the proof. We have assumed that $\xi \geq \mu$ and from the condition of Theorem 3, we have $\xi \leq 2^n/8q_m$. By assuming $\mu \leq 2^n/8q_m$ (otherwise the bound becomes vacuously true) we choose $\xi = 2^n/8q_m$. Hence, the result follows by applying Eqn. (2), Lemma 2, Eqn. (26) and $\xi = 2^n/8q_m$. □

6.3 Security Bound using the Coefficients-H Technique

We instantiate the underlying hash function of nEHtM by a truncated n-bit PolyHash function that truncates the first bit of the PolyHash output which is $2\ell/2^n$-axu hash function [14], where ℓ is the maximum number of message blocks. Therefore, from Lemma 2, Eqn. (22) and the inequality $\sum_{i=1}^{k}\binom{W_i}{2} \leq \xi q_m$, we obtain the following bound using the coefficients-H technique.

$$\delta_{\text{hc}} \leq \frac{q_m + 2q_v}{2^n} + \frac{q_m^2\ell}{2^n\xi} + \frac{(2q_m + q_v)2\ell\mu}{2^n} + \frac{2q_v\ell}{2^n} + \frac{24q_m^3\xi}{2^{2n}}. \tag{27}$$

We choose the optimal value of ξ such that the right hand side of the Eqn. (27) gets maximized. To obtain such a value of ξ, we must have $\frac{q_m^2\ell}{2^n\xi} = \frac{24q_m^3\xi}{2^{2n}}$. By solving the equality for ξ, we obtain $\xi_{\text{opt}} = \left(\frac{\ell 2^n}{24q_m}\right)^{\frac{1}{2}}$. Plugging-in this optimal value of ξ_{opt} into Eqn. (27) gives

$$\delta_{\text{hc}} \leq \frac{q_m + 2q_v}{2^n} + \frac{(2q_m + q_v)2\ell\mu}{2^n} + \frac{2q_v\ell}{2^n} + 10\left(\frac{q_m^5\ell}{2^{3n}}\right)^{\frac{1}{2}}.$$

The above bound holds true as long as $q \leq 2^{3n/5}/\ell^{1/5} \approx O(2^{3n/5})$, which is weaker than the bound $O(2^{2n/3})$ that we obtained using the expectation method.

7 Proof of Theorem 2

In this section we prove Theorem 2. Instead of separately proving the privacy and the authenticity result of the construction, we bound the distinguishing advantage of the two random systems: (i) the pair of oracles $(\mathsf{CWC+.Enc}, \mathsf{CWC+.Dec})$ for a random permutation Π, which is called the real system or the real world and (ii) the pair of oracles $(\mathsf{Rand}, \mathsf{Rej})$, which is called the ideal system or the ideal world. The privacy and authenticity bounds of $\mathsf{CWC+}$ then follow as a simple corollary of this result. We prove the following information theoretic bound of $\mathsf{CWC+}$.

$$\delta^* \leq \frac{97\sigma^3\ell}{2^{2n}} + \frac{5\sigma}{2^n} + \frac{\sigma\ell}{2^n} + \frac{8\sigma^3}{2^{2n}} + \frac{2q_d}{2^\rho}\left(1 + \frac{\ell}{2^{n-\rho}}\right) + \frac{(2q_e + q_d)2\ell\mu}{2^n} + \left(\frac{5\sigma\ell\mu}{2^n}\right)^2, \tag{28}$$

where δ^* is the maximum advantage in distinguishing the real world from the ideal world and we assume $q_e\ell \approx \sigma$, $\sigma \leq 2^n/48$. Due to limitations in space, we provide here only a sketch of the proof, and details may be found in Supplementary Sect. D.

Description of the ideal world. We begin with the assumption that all the queried messages of an adversary are of length multiple of n and the number of blocks of i-th message is l_i. Now, we consider a deterministic distinguisher A that interacts either with the real world or with the ideal world. Rej simply rejects all the verification attempts of A whereas Rand, on the i-th encryption query (N_i, M_i, A_i) works as shown in Fig. 7.1.

```
Algorithm Rand$(N_i, A_i, M_i)$

if $N_i \in \mathcal{D}$, let $N_i = N$
  if $l_i = l_N$, then $S_i \leftarrow \mathcal{L}(N)$
  if $l_i < l_N$, then $S_i \leftarrow \mathcal{L}(N)[1, nl_i]$
  if $l_i > l_N$, then
    $R \leftarrow\!\!\$\, (\{0,1\}^n)^{l_i - l_N}, S_i \leftarrow \mathcal{L}(N)\|R$
    $l_N = l_i$
else
  $S_i \leftarrow\!\!\$\, (\{0,1\}^n)^{l_i}$, $\mathcal{L}(N_i) \leftarrow S_i, l_{N_i} = l_i$
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{N_i\}$
$\widetilde{T}_i \leftarrow\!\!\$\, \{0,1\}^n$; $T_i \leftarrow \mathsf{chop}_\rho(\widetilde{T}_i)$
return $(S_i, T_i)$
```

Fig. 7.1. Random oracle for the ideal world. l_N denotes the updated number of keystream blocks for nonce N and $\mathcal{L}(N)$ denotes the updated keystream blocks for nonce N of length l_N. $\mathcal{D}$ denotes the domain of the nonce. chop_ρ is a function that truncates the last $n - \rho$ bits of its input.

Attack Transcript. Let D be a fixed non-trivial computationally unbounded deterministic distinguisher that interacts with either the real world or the ideal world, making at most q_e queries to the left (encryption) oracle with at most μ faulty nonces and at most q_d queries to its right (decryption) oracle, and returning a single bit.

Let $\tau_e \triangleq \{(N_1, M_1, A_1, S_1, T_1), \ldots, (N_{q_e}, M_{q_e}, A_{q_e}, S_{q_e}, T_{q_e})\}$ be the list of encryption queries and $\tau_d \triangleq \{(N_1', A_1', C_1', T_1', Z_1), \ldots, (N_{q_d}', A_{q_d}', C_{q_d}', T_{q_d}', Z_{q_d})\}$ be the list of decryption queries, where $Z_i = M_i \cup \{\bot\}$. Note that the encryption oracle in both the worlds releases the keystream as it determines the ciphertext uniquely. For convenience, we reveal the hash key K_h, which is $\mathsf{E}_K(\mathbf{0})$, if D interacts with the real world or a uniform random element from $\{0,1\}^n$, if D interacts with the ideal world, and also the n-bit tag (without truncating) i.e., $\mathbf{T} \triangleq (\widetilde{T}_1, \ldots, \widetilde{T}_{q_e})$ to the distinguisher after it made all its queries and obtains corresponding responses but before it output its decision and thus the extended query transcript of the attack is $\tau' = (\tau, K_h, \widetilde{\mathbf{T}})$, which is called the *extended transcript.*

Bad Transcripts. Recall that N_i is a $3n/4$-bit string. We denote $0\|N_i\|0^{n/4-1}$ as $\widehat{N}_i$ and $1\|X_i$ as $\widehat{X}_i$, where $X_i \stackrel{\Delta}{=} N_i\|0^{n/4-1} \oplus \mathsf{Poly}_{K_h}(M_i)$. Moreover, $S_i[j]$ denotes the j-th keystream block for i-th message. With these notations, we define the bad transcript as follows: a transcript $\tau' = (\tau_e, \tau_d, K_h, \widetilde{\mathbf{T}})$ is called **bad** if any one of the following holds:

- B.1 : $\exists\ i \in [q_e]$ and $j \in [l_i]$ such that $S_i[j] = K_h$.
- B.2 : $\exists\ i \in [q_e]$ and $j \in [l_i]$ such that $S_i[j] = \mathbf{0}$.
- B.3 : $\exists\ i \in [q_e]$ and $j, j' \in [l_i]$ such that $S_i[j] = S_i[j']$.
- B.4 : $\exists\ i \in [q_e]$ such that $\widetilde{T}_i = \mathbf{0}$.
- B.5 : $\exists i \neq j, j \neq k$ such that $\widehat{N}_i = \widehat{N}_j$ and $\widehat{X}_j = \widehat{X}_k$.
- B.6 : $\{i_1, \ldots, i_{\xi+1}\} \subseteq [q_e]$ such that $\widehat{X}_{i_1} = \widehat{X}_{i_2} = \ldots = \widehat{X}_{i_{\xi+1}}$ for some parameter $\xi \geq \mu$.
- B.7 $\exists\ a \in [q_d]$, $\exists\ i \in [q_e]$ such that $\widehat{N}_i = \widehat{N'}_a$, $\widehat{X}_i = \widehat{X'}_a$ and $\widetilde{T}_i = T'_a$.

Θ_{bad} (resp. Θ_{good}) denotes the set of bad (resp. good) transcripts. Moreover, X_{re} and X_{id} denotes the probability distribution of realizing an extended transcript τ' in the real and the ideal world respectively. We bound the probability of bad transcripts in the ideal world as follows.

Lemma 3. *Let* X_{id} *and* Θ_{bad} *be defined as above. Then*

$$\Pr[X_{\text{id}} \in \Theta_{\text{bad}}] \leq \epsilon_{\text{bad}} = \frac{2\sigma}{2^n} + \frac{q_e \ell^2}{2^n} + \frac{q_e}{2^n} + \frac{q_e^2 \ell}{\xi 2^n} + \frac{(2q_e + q_d)2\ell\mu}{2^n} + \frac{2q_d\ell}{2^n}.$$

Proof of the lemma can be found in Supplementary Sect. D.

Good Transcripts. Let $\tau' = (\tau_e, \tau_d, K_h, \widetilde{\mathbf{T}})$ be a good transcript. Since in the ideal world the encryption oracle is perfectly random and the decryption oracle always rejects, one simply has

$$\Pr[X_{\text{id}} = \tau'] = \frac{1}{2^n} \cdot \prod_{t=1}^{r} \frac{1}{2^{nl_t}} \cdot \frac{1}{2^{nq_e}} \tag{29}$$

where r is the number of groups of nonces and l_t be the updated number of generated keystream blocks for group t.

Real Interpolation Probability. To bound the probability of getting τ' in the real world from below, we model the system of equations and non-equations into the graph theoretic setting to obtain the graph $G_{\tau'}$, where we have $\sigma + q_e$ equations and $2^{n-\rho}q_d$ non-equations. Similar to the analysis of good transcripts in the proof of Theorem 1, one can argue that as τ' is good, $G_{\tau'}$ is valid (i.e., it satisfies NC, NPL and NCL conditions). Thus, we assume $\xi \leq 2^n/8\sigma\ell$ (from the condition of Theorem 3), which allows us to apply Theorem 3 with $c = 1$,

$\sigma_{i-1} \leq \sigma_k \leq 2\sigma$ and $\alpha \leq \sigma$ and then dividing by Eqn. (29) we have,

$$\frac{\Pr[X_{\mathrm{re}} = \tau']}{\Pr[X_{\mathrm{id}} = \tau']} \geq 1 - \underbrace{\left(\sum_{i=1}^{k} \frac{24\sigma^2 \binom{W'_i}{2}}{2^{2n}} + \frac{2q_d}{2^\rho} + \frac{2\sigma}{2^n}\right)}_{\phi(\tau')}, \tag{30}$$

where k is the number of components of $G_{\tau'}$ and W'_i denotes the size of the i-th component. Note that there are $2^{n-\rho}q_d$ non-equations as the adversary forges with ρ bit tags T'_a and there are $2^{n-\rho}$ tags $\widetilde{T}$s whose first ρ bits match with T'_a. Moreover, we consider $c = 1$ due to the fact that we choose elements from the set $\{0,1\}^n$ excluding the hash key.

Finalizing the proof. We calculate the expectation of $\phi(\tau')$ as follows:

$$\mathbf{E}[\phi(X_{\mathrm{id}})] = \left(\frac{2q_d}{2^\rho} + \frac{2\sigma}{2^n} + \frac{24\sigma^2}{2^{2n}}\mathbf{E}\left[\sum_{i=1}^{k}\binom{W'_i}{2}\right]\right). \tag{31}$$

It is easy to see that $\binom{W'_i}{2} \leq \binom{W_i}{2}\binom{2\ell}{2}$, where W_i is defined in the proof of Theorem 1. Therefore from Eqn. (24) and Eqn. (25),

$$\mathbf{E}\left[\sum_{i=1}^{k}\binom{W'_i}{2}\right] \leq \frac{2q_e^2\ell^3}{2^n} + \mu^2\ell^2 + 4q_e\ell^2, \tag{32}$$

where the almost xor universal probability of the truncated PolyHash is at most $2\ell/2^n$. Finally, from Eqn. (31) and Eqn. (32) we obtain

$$\mathbf{E}[\phi(X_{\mathrm{id}})] \leq \left(\frac{2q_d}{2^\rho} + \frac{2\sigma}{2^n} + \frac{48\sigma^4\ell}{2^{3n}} + \left(\frac{5\sigma\ell\mu}{2^n}\right)^2 + \frac{96\sigma^3\ell}{2^{2n}}\right), \tag{33}$$

where we assume that $\ell q_e \approx \sigma$, the total number of message blocks queried.

Finalization. We have assumed that $\xi \geq \mu$ and from the condition of Theorem 3, we have $\xi \leq 2^n/8\sigma\ell$. By assuming $\mu \leq 2^n/8\sigma\ell$ (otherwise the bound becomes vacuously true) we choose $\xi = 2^n/8\sigma\ell$. Hence, the bound stated in Eqn. (28) follows by applying Eqn. (2), Lemma 3, Eqn. (33), $\xi = 2^n/8\sigma\ell$ and $\sigma \leq 2^n/48$. □

Concluding the proof of theorem 2. The privacy bound of CWC+ is derived from Eqn. (28) by setting $\mu = 0$ and the bound stated in Eqn. (28) is itself the authenticity bound of CWC+.

Acknowledgements. Authors would like to thank to all the anonymous reviewers of Eurocrypt, 2019.

References

1. Caesar: Competition for authenticated encryption: Security, applicability, and robustness.
2. A.Joux. Comments on the draft gcm specification authentication fail- ures in nist version of gcm.
3. Kazumaro Aoki and Kan Yasuda. The security and performance of "gcm" when short multiplications are used instead. In *Information Security and Cryptology - 8th International Conference, Inscrypt 2012, Beijing, China, November 28-30, 2012, Revised Selected Papers*, pages 225–245, 2012.
4. Tomer Ashur, Orr Dunkelman, and Atul Luykx. Boosting authenticated encryption robustness with minimal modifications. In *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part III*, pages 3–33, 2017.
5. Mihir Bellare and Chanathip Namprempre. Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In *Advances in Cryptology - ASIACRYPT 2000, 6th International Conference on the Theory and Application of Cryptology and Information Security, Kyoto, Japan, December 3-7, 2000, Proceedings*, pages 531–545, 2000.
6. Mihir Bellare and Björn Tackmann. The multi-user security of authenticated encryption: AES-GCM in TLS 1.3. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, pages 247–276, 2016.
7. Daniel J. Bernstein. The poly1305-aes message-authentication code. In *Fast Software Encryption: 12th International Workshop, FSE 2005, Paris, France, February 21-23, 2005, Revised Selected Papers*, pages 32–49, 2005.
8. Srimanta Bhattacharya and Mridul Nandi. Revisiting variable output length XOR pseudorandom function. *IACR Trans. Symmetric Cryptol.*, 2018(1):314–335, 2018.
9. Hanno Böck, Aaron Zauner, Sean Devlin, Juraj Somorovsky, and Philipp Jovanovic. Nonce-disrespecting adversaries: Practical forgery attacks on GCM in TLS. In *10th USENIX Workshop on Offensive Technologies, WOOT 16, Austin, TX, USA, August 8-9, 2016.*, 2016.
10. Priyanka Bose, Viet Tung Hoang, and Stefano Tessaro. Revisiting AES-GCM-SIV: multi-user security, faster key derivation, and better bounds. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part I*, pages 468–499, 2018.
11. B.Smith. Pull request: Removing the aead explicit iv. mail to ietf tls working group. 2015.
12. Ran Canetti and Hugo Krawczyk. Analysis of key-exchange protocols and their use for building secure channels. In *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceeding*, pages 453–474, 2001.
13. 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.
14. Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Goutam Paul. Double-block hash-then-sum: A paradigm for constructing bbb secure prf. *IACR Transactions on Symmetric Cryptology*, 2018(3):36–92, 2018.

15. 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.
16. Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Kan Yasuda. Encrypt or decrypt? to make a single-key beyond birthday secure nonce-based MAC. In *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part I*, pages 631–661, 2018.
17. Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Kan Yasuda. Encrypt or decrypt? to make a single-key beyond birthday secure nonce-based mac. Cryptology ePrint Archive, Report 2018/500, 2018.
18. Avijit Dutta, Ashwin Jha, and Mridul Nandi. Tight security analysis of ehtm MAC. *IACR Trans. Symmetric Cryptol.*, 2017(3):130–150, 2017.
19. Viet Tung Hoang and Stefano Tessaro. Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, pages 3–32, 2016.
20. Viet Tung Hoang and Stefano Tessaro. The multi-user security of double encryption. In *Advances in Cryptology - EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 - May 4, 2017, Proceedings, Part II*, pages 381–411, 2017.
21. Tetsu Iwata. New blockcipher modes of operation with beyond the birthday bound security. In *Fast Software Encryption, 13th International Workshop, FSE 2006, Graz, Austria, March 15-17, 2006, Revised Selected Papers*, pages 310–327, 2006.
22. Tetsu Iwata. Authenticated encryption mode for beyond the birthday bound security. In *Progress in Cryptology - AFRICACRYPT 2008, First International Conference on Cryptology in Africa, Casablanca, Morocco, June 11-14, 2008. Proceedings*, pages 125–142, 2008.
23. Tadayoshi Kohno, John Viega, and Doug Whiting. CWC: A high-performance conventional authenticated encryption mode. In *Fast Software Encryption, 11th International Workshop, FSE 2004, Delhi, India, February 5-7, 2004, Revised Papers*, pages 408–426, 2004.
24. Will Landecker, Thomas Shrimpton, and R. Seth Terashima. Tweakable blockciphers with beyond birthday-bound security. In *Advances in Cryptology - CRYPTO 2012 - 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings*, pages 14–30, 2012.
25. David A. McGrew and John Viega. The security and performance of the galois/counter mode (GCM) of operation. In *Progress in Cryptology - INDOCRYPT 2004, 5th International Conference on Cryptology in India, Chennai, India, December 20-22, 2004, Proceedings*, pages 343–355, 2004.
26. Bart Mennink and Samuel Neves. Encrypted davies-meyer and its dual: Towards optimal security using mirror theory. Cryptology ePrint Archive, Report 2017/473, 2017.
27. Bart Mennink and Samuel Neves. Encrypted davies-meyer and its dual: Towards optimal security using mirror theory. In *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part III*, pages 556–583, 2017.
28. Kazuhiko Minematsu. How to thwart birthday attacks against macs via small randomness. In *Fast Software Encryption, FSE 2010*, pages 230–249, 2010.

29. Kazuhiko Minematsu and Tetsu Iwata. Building blockcipher from tweakable blockcipher: Extending FSE 2009 proposal. In *Cryptography and Coding - 13th IMA International Conference, IMACC 2011, Oxford, UK, December 12-15, 2011. Proceedings*, pages 391–412, 2011.
30. Chanathip Namprempre, Phillip Rogaway, and Thomas Shrimpton. Reconsidering generic composition. In *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, pages 257–274, 2014.
31. Mridul Nandi. Birthday attack on dual ewcdm. Cryptology ePrint Archive, Report 2017/579, 2017. https://eprint.iacr.org/2017/579.
32. Jacques Patarin. The "coefficients h" technique. In *Selected Areas in Cryptography, 15th International Workshop, SAC 2008, Sackville, New Brunswick, Canada, August 14-15, Revised Selected Papers*, pages 328–345, 2008.
33. 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.
34. 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.
35. Jacques Patarin. Mirror theory and cryptography. *IACR Cryptology ePrint Archive*, 2016:702, 2016.
36. Thomas Peyrin and Yannick Seurin. Counter-in-tweak: Authenticated encryption modes for tweakable block ciphers. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, pages 33–63, 2016.
37. Phillip Rogaway. Efficient instantiations of tweakable blockciphers and refinements to modes OCB and PMAC. In *Advances in Cryptology - ASIACRYPT 2004, 10th International Conference on the Theory and Application of Cryptology and Information Security, Jeju Island, Korea, December 5-9, 2004, Proceedings*, pages 16–31, 2004.
38. Victor Shoup. On fast and provably secure message authentication based on universal hashing. In *Advances in Cryptology - CRYPTO '96, 16th Annual International Cryptology Conference, Santa Barbara, California, USA, August 18-22, 1996, Proceedings*, pages 313–328, 1996.
39. Mark N. Wegman and Larry Carter. New hash functions and their use in authentication and set equality. *J. Comput. Syst. Sci.*, 22(3):265–279, 1981.
40. Ping Zhang, Honggang Hu, and Qian Yuan. Close to optimally secure variants of GCM. *Security and Communication Networks*, 2018:9715947:1–9715947:12, 2018.

Supplementary Materials

A Nonce Misuse Attack on nEHtM

In the following we discuss the birthday bound forging attack on nEHtM when the number of faulty query is roughly $2^{n/2}$. As stated before that the underlying idea of the attack is to form an alternating cycle of length 4 in the input of the block cipher. For this an adversary A makes two sets of $2^{n/2}$ MAC queries;

one with message M and another with message $M'(\neq M)$ and then finds four queries such that the sum of their tag becomes zero. A mounts the attack in the following two phases. (a) In the first phase it finds out the quadruple that makes the tag sum zero. (b) In the second phase it forges the MAC. The algorithmic description of the attack is shown in part (b) of Fig. A.1.

First phase of the attack:

1. A makes $q^* = 2^{n/2}$ MAC queries with distinct nonces but same message M, i.e., $(N_1, M) \mapsto T_1, (N_2, M) \mapsto T_2, \ldots, (N_{q^*}, M) \mapsto T_{q^*}$, where $T_1, T_2, \ldots, T_{q^*}$ are the tags.
2. A makes another $2^{n/2}$ MAC queries with same nonces but different message M', i.e., $(N_{q^*+1}, M') \mapsto T_{q^*+1}, \ldots, (N_{2q^*}, M') \mapsto T_{2q^*}$, where $N_{q^*+i} = N_i$ for all $i \in [q^*]$ and $T_{q^*+1}, T_{q^*+2}, \ldots, T_{2q^*}$ are the tags.
3. A finds two distinct query indices $i, j \in [q^*]$ such that $T_i \oplus T_j \oplus T_{q^*+i} \oplus T_{q^*+j} = \mathbf{0}$.

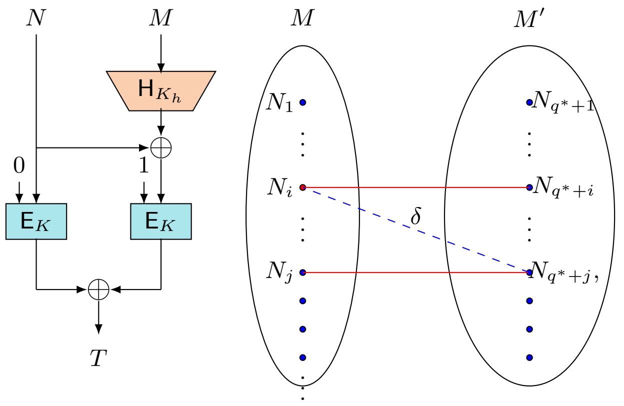

Fig. A.1. (a) Left part is the domain separation variant of *Nonce based Enhanced Hash-then-Mask* with an n-bit keyed hash function H_{K_h} and an n-bit block cipher E_K; (b) Right part is the forging attack on the construction.

Note that, the event $\mathsf{CollT} \triangleq T_i \oplus T_j \oplus T_{q^*+i} \oplus T_{q^*+j} = \mathbf{0}$ can take place either because of (i) the collision of the hash i.e., $\mathsf{H}_K(M) \oplus \mathsf{H}_K(M') = N_i \oplus N_j$ or (ii) due to the random output of the underlying permutation Π. Probability of occurring the second case is extremely low (we call it as **false positive**) and therefore, when CollT takes place, we can assume with high probability that the hash value collides. As a result, A obtains the hash difference $N_i \oplus N_j$.

Second phase of the attack:

A chooses two distinct nonces $N_{2q^*+1}, N_{2q^*+2} \notin \{N_1, \ldots, N_{q^*}\}$ such that $N_{2q^*+1} \oplus N_{2q^*+2} = N_i \oplus N_j$ and makes queries $(N_{2q^*+1}, M) \mapsto T_{2q^*+1}, (N_{2q^*+1}, M') \mapsto$

T_{2q^*+2} and $(N_{2q^*+2}, M) \mapsto T_{2q^*+3}$. This allows A to forge with $(N_{2q^*+2}, M', T_{2q^*+1} \oplus T_{2q^*+2} \oplus T_{2q^*+3})$.
Note that, the step (3) holds with probability $(q^*)^2/2^n$. Therefore, the above attack holds for $q \approx 2^{n/2+1}$ and $\mu = 2^{n/2}$, but $\nu = 2$.

B Proof of Theorem 3

Before we prove the theorem, we remind the reader that we can represent an extended system of equations and non-equations with a simple undirected edge-labelled graph G where the edge set of the graph is partitioned into two disjoint sets $\mathcal{S}$ and $\mathcal{S}'$ and $\mathcal{V} = \{Y_1, \dots, Y_\alpha\}$ such that the following happens:

$$Y_i \oplus Y_j = \lambda_{ij} \ \forall \ \{Y_i, Y_j\} \in \mathcal{S}, \tag{34}$$

$$Y_i \oplus Y_j \neq \lambda'_{ij} \ \forall \ \{Y_i, Y_j\} \in \mathcal{S}', \tag{35}$$

where λ_{ij} is the label of the edge $\{Y_i, Y_j\} \in \mathcal{S}$ and λ'_{ij} is the label of the edge $\{Y_i, Y_j\} \in \mathcal{S}'$. We are interested to find a good lower bound on the number of injective solutions to the system of linear equations and non-equations induced by the graph G. We said that G is valid if it satisfies NC, NPL and NCL properties. We once again recall Theorem 3.

Theorem 4. *Let $G = (\mathcal{V}, \mathcal{S} \sqcup \mathcal{S}', \mathcal{L})$ be a good graph with α vertices such that $|\mathcal{S}| = q_m, |\mathcal{S}'| = q_v$. Let $\mathsf{comp}(G^=) = (\mathcal{C}_1, \dots, \mathcal{C}_k)$ and $|\mathcal{C}_i| = w_i$, $\sigma_i = (w_1 + \dots + w_i)$. Then, the total number of injective solutions, chosen from a set $\mathcal{Z}$ of size $2^n - c$, for some $c \geq 0$, for the induced system of equations and non-equations $\mathcal{E}_G$ is at least:*

$$\frac{(2^n)_\alpha}{2^{nq}} \left(1 - \sum_{i=1}^{k} \frac{6\sigma_{i-1}^2 \binom{w_i}{2}}{2^{2n}} - \frac{2(q_v + c\alpha)}{2^n}\right),$$

provided $\sigma_k w_{\max} \leq 2^n/4$.

As a warm up, the reader may first consider the same problem with only a system of affine equations. In specific, we prove the following result.

Lemma 4. *Let $G = (\mathcal{V}, \mathcal{S}, \mathcal{L})$ be a simple edge-labelled undirected acyclic graph that satisfies the NPL condition. Let $\mathsf{COMP}(G) = (\mathcal{C}_1, \dots, \mathcal{C}_k)$ be the set of component of G such that $|\mathcal{C}_i| = w_i$, for each $i = 1, \dots, k$ and let the number of edges in G is q. We denote $\sigma_i = (w_1 + \dots, w_i)$, to be the number of vertices upto the first i components of G with $\sigma_0 = 0$. Then, the total number of injective solutions for the induced system of equations $\mathcal{E}_G$, denoted by h_α, is at least*

$$h_\alpha \geq \frac{(2^n)_\alpha}{2^{nq}} \left(1 - \sum_{i=1}^{k} \frac{6\sigma_{i-1}^2 \binom{w_i}{2}}{2^{2n}}\right),$$

provided $\sigma_k w_{\max} \leq 2^n/4$, where $w_{\max} = \max\{w_1, \dots, w_k\}$.

Proof. Consider the first component $\mathcal{C}_1$ of the graph G. Let $Y_{i_1} \in \mathcal{V}$ be any arbitrary vertex of $\mathcal{C}_1$. There are 2^n choices for assigning values to the variable Y_{i_1}. Let the assigned value to Y_{i_1} be y_{i_1}. Now, for any other variable Y_{i_2} of $\mathcal{C}_1$, we consider the path P from Y_{i_1} to Y_{i_2} and assign the value $y_{i_1} \oplus \mathcal{L}(P)$ to the variable Y_{i_2}. Let this value be y_{i_2}. Note that, the path is unique as the graph is acyclic and due to the NPL property, $\mathcal{L}(P) \neq \mathbf{0}$ and hence $y_{i_1} \neq y_{i_2}$. Therefore, assigned values to all the variables in $\mathcal{C}_1$ is different from y_i.

Now, we argue that if Y_j and Y_k are any two arbitrary variables in $\mathcal{C}_1$, then the assigned values to them must be distinct. Suppose, P_j and P_k are the paths from the vertex Y_{i_1} to Y_j and Y_k respectively (the path should exists and it is unique as the component is connected and contains no cycle). Let P be the common prefix (which may be empty) of P_j and P_k. Therefore, we can write $P_j = P \| P'_j,\ P_k = P \| P'_k$. Note that $P'_j \| P'_k$ is itself is the path from vertex Y_j to Y_k. Now, from the definition $y_j = y_{i_1} \oplus \mathcal{L}(P_j)$ and $y_k = y_{i_1} \oplus \mathcal{L}(P_k)$, where y_j and y_k are the assigned values to the variable Y_j and Y_k respectively and hence,

$$y_j \oplus y_k = \mathcal{L}(P_j) \oplus \mathcal{L}(P_k) = \mathcal{L}(P'_j) \oplus \mathcal{L}(P'_k) = \mathcal{L}(P'_j \| P'_k) \neq \mathbf{0},$$

where the last equality holds due to the NPL condition. It is also straightforward to verify that for all edges $\{j,k\} \in \mathcal{S}$, $y_j \oplus y_k = \lambda_{jk}$. Therefore, y_{i_1} sets the solution uniquely to all the variables in $\mathcal{C}_1$. Let $(y_{i_1}, \ldots, y_{i_{w_1}})$ denotes one such possible solution, where each element of the tuple is pairwise distinct and hence the tuple is an injective solution to all the variables in the first component. Once such a value is fixed for Y_{i_1}, we consider the second component.

We do a similar calculation for the second component $\mathcal{C}_2$. Let $Y_{i_{w_1+1}} \in \mathcal{V}$ be a variable in $\mathcal{C}_2$. For any *valid solution* $y_{i_{w_1+1}}$ for $Y_{i_{w_1+1}}$, we set $y_{i_{w_1+1}} \oplus \mathcal{L}(P)$ as a solution to the variable $Y_j \in \mathcal{V}$, where Y_j be any arbitrary vertex in $\mathcal{C}_2$ and P is the unique path from $Y_{i_{w_1+1}}$ to Y_j. Therefore, $y_{i_{w_1+1}}$ actually uniquely determines the values of the remaining $w_2 - 1$ variables. Now, if $y_{i_{w_1+1}}$ is a valid solution to $Y_{i_{w_1+1}}$, then

- $y_{i_{w_1+1}}$ must be distinct from $(y_{i_1}, \ldots, y_{i_{w_1}})$ values which are already been assigned to the variables in $\mathcal{C}_1$.
- $y_{i_{w_1+1}}$ must be distinct from $(y_{i_1} \oplus \mathcal{L}(P_j), \ldots, y_{i_{w_1}} \oplus \mathcal{L}(P_j))$ for all possible paths P_j, the path from the vertex $Y_{i_{w_1+1}}$ to any other vertex Y_j in $\mathcal{C}_2$.

Thus, at most $w_1 w_2$ values get discarded for assignment to the vertex $Y_{i_{w_1+1}}$. Thus, there are at least $(2^n - w_1 w_2)$ choices for assigning values to the vertex $Y_{i_{w_1+1}}$ and hence the number of injective solutions to exist for the second component.

In general, for the i-th component, once the injective solution is fixed for previous $i-1$ components, there are at least $(2^n - w_1 w_i - \cdots - w_{i-1} w_i) = (2^n - \sigma_{i-1} w_i)$ ways for an injective solution to exist for the i-th component, when the vertices of the first $i-1$ components have already been assigned values. Hence, the total number of possible injective solutions for the induced system of equations is at least

$$h_\alpha \geq \prod_{i=1}^{k}\Big(2^n - \sigma_{i-1}w_i\Big) \tag{36}$$

Finalizing the proof of Lemma. By doing a simple algebra on Eqn. (36) we have,

$$h_\alpha \frac{2^{nq}}{(2^n)_\alpha} \geq \frac{2^{nq}}{(2^n)_\alpha}\prod_{i=1}^{k}(2^n - \sigma_{i-1}w_i) = \prod_{i=1}^{k}\frac{(2^n - \sigma_{i-1}w_i)2^{n(w_i-1)}}{(2^n - \sigma_{i-1})_{w_i}}, \tag{37}$$

where q and α is the number of edges and vertices of G respectively. Now, we have

$$\begin{aligned}(2^n - \sigma_{i-1})_{w_i} &\overset{(1)}{\leq} 2^{nw_i} - 2^{n(w_i-1)}\left(\sigma_{i-1}w_i + \binom{w_i}{2}\right) \\ &\quad + 2^{n(w_i-2)}\underbrace{\left(\binom{w_i}{2}\sigma_{i-1}^2 + \binom{w_i}{2}(w_i-1)\sigma_{i-1} + \binom{w_i}{2}\frac{(w_i-2)(3w_i-1)}{12}\right)}_{A_i}.\end{aligned}$$

Plugging-in the inequality (1) into Eqn. (37) gives

$$\begin{aligned}h_\alpha \frac{2^{nq}}{(2^n)_\alpha} &\geq \prod_{i=1}^{k}\left(1 + \frac{2^{n(w_i-1)}\cdot\binom{w_i}{2} - 2^{n(w_i-2)}\cdot A_i}{2^{nw_i} - 2^{n(w_i-1)}(\sigma_{i-1}w_i + \binom{w_i}{2}) + 2^{n(w_i-2)}A_i}\right) \\ &\geq \prod_{i=1}^{k}\left(1 - \frac{A_i}{2^{2n} - 2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) + A_i}\right) \\ &\overset{(2)}{\geq} \prod_{i=1}^{k}\left(1 - \frac{2A_i}{2^{2n}}\right) \overset{(3)}{\geq} \left(1 - \sum_{i=1}^{k}\frac{6\sigma_{i-1}^2\binom{w_i}{2}}{2^{2n}}\right),\end{aligned}$$

where (2) follows from the fact that $2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) - A_i \leq 2^{2n}/2$, which holds to true when $\sigma_k w_{\max} \leq 2^n/4$ and (3) holds to true due to the fact that $A_i \leq 3\sigma_{i-1}^2\binom{w_i}{2}$. □

B.1 Proof of Theorem 3

To prove Theorem 3, we first state and prove the following lemma. Proof of Theorem 3 will then be directly followed from Lemma 5 (proven below) and the constraint $\sigma_k w_{\max} \leq 2^n/4$, where $w_{\max}$ denotes the maximum component size of the graph $G^=$. Recall that, q_v denotes the number of non-equation edges in graph G.

The proof of the following lemma will be similar to that of Lemma 4, the only difference is that we need to incorporate the non-equation edges along with the equation edges and that makes a difference in the counting.

Notation. For the purpose of proving the following lemma, we recall the following notation: a *blue dashed edge* represents a non-equation edge and hence belongs to the set $\mathcal{S}'$ and a *red continuous edge* represents a equation edge and hence belongs to the set $\mathcal{S}$. Moreover, $G^{=}(\mathcal{V})$ denotes the set of all vertices of the subgraph $G^{=}$ and the graph G is said to be valid if it satisfies the *"no-cycle"*, *"NPL"* and *"NCL"* properties.

Lemma 5. *Let $G = (\mathcal{V}, \mathcal{S} \sqcup \mathcal{S}', \mathcal{L})$ be a valid graph such that $\mathcal{V} = \{Y_1, \ldots, Y_\alpha\}$ and $|\mathcal{S}| = q_m, |\mathcal{S}'| = q_v$. Let $\mathsf{COMP}(G^{=}) = (\mathcal{C}_1, \ldots, \mathcal{C}_k)$ be the set of components of $G^{=}$ such that $|\mathcal{C}_i| = w_i$, for each $i = 1, \ldots, k$. For every $i \neq j \in [k]$, suppose there are $\tilde{w}_{ij}$ edges from $\mathcal{S}'$ connecting vertices of the i-th and j-th components of $G^{=}$ and $\sigma_i = (w_1 + \ldots + w_i)$ denotes the number of vertices upto the first i components of $G^{=}$ with $\sigma_0 = 0$. Moreover, let $|\mathcal{V} \setminus G^{=}(\mathcal{V})| = k'$ and for any vertex $v_i \in \mathcal{V} \setminus G^{=}(\mathcal{V})$, there are w'_i blue dashed edges incident on v_i. Then, the total number is injective solutions, chosen from a set $\mathcal{Z}$ of size $2^n - c$, for some $c \geq 0$, for the induced system of equations and non-equations $\mathcal{E}_G$, denoted by h_α, is at least*

$$h_\alpha \geq \prod_{i=1}^{k} \left(2^n - \sigma_{i-1} w_i - \sum_{j=1}^{i-1} \tilde{w}_{ij} - c w_i\right) \cdot \prod_{i \in [k']} \left(2^n - \sigma_k - i - w'_i\right). \tag{38}$$

Proof. There are clearly $(2^n - cw_1)$ ways to assign values to any one of the vertices of the first component $\mathcal{C}_1$ of $G^{=}$ which uniquely determines the assigned values to the rest of the variables in $\mathcal{C}_1$, as argued in the proof of Lemma 4. Thus, there are $(2^n - cw_1)$ ways for an injective solution to exist for the first component. Once such a solution is fixed for the first component, we consider the second component.

We consider any arbitary vertex in the second component $\mathcal{C}_2$ of $G^{=}$. Let $Y_{i_{w_1+1}} \in \mathcal{V}$ be a variable in $\mathcal{C}_2$ and we have argued in the proof of Lemma 4 that for any valid solution $y_{i_{w_1+1}}$ for $Y_{i_{w_1+1}}$, $Y_{i_{w_1+1}}$ should not take $w_1 w_2$ values. Additionally, as there are $\tilde{w}_{12}$ blue dashed edges connecting the component $\mathcal{C}_1$ and $\mathcal{C}_2$, there are $\tilde{w}_{12}$ paths from the vertex $Y_{i_{w_1+1}}$ to the vertices of the component $\mathcal{C}_1$. Moreover, $y_{i_{w_1+1}}$ cannot take additional cw_2 values. As a result, if $y_{i_{w_1+1}}$ is a valid solution to the variable $Y_{i_{w_1+1}}$, then

$y_{i_{w_1+1}}$ *should not take $\tilde{w}_{12}$ values that violates the non-equality conditions of $\tilde{w}_{12}$ blue dashed edges and also cw_2 values.*

Thus, there are at most $w_1 w_2 + \tilde{w}_{12} + cw_2$ values get discarded for assignment to the vertex $Y_{i_{w_1+1}}$ and as a result there are at least $(2^n - w_1 w_2 - \tilde{w}_{1,2} - cw_2)$ valid choices for $Y_{i_{w_1+1}}$. Once a valid value is assigned to the variable $Y_{i_{w_1+1}}$, remaining variables in the second component will be assigned uniquely. Thus, there are $(2^n - w_1 w_2 - \tilde{w}_{12} - cw_2)$ ways for an injective solution to exist for the second component.

In general, for the i-th component, once the injective solution is fixed for previous $i - 1$ components, there are at least $(2^n - \sigma_{i-1} w_i - \tilde{w}_{i1} - \ldots - \tilde{w}_{i,i-1} - cw_i)$ ways for an injective solution to exist for the i-th component, when the vertices of the first $i-1$ components have already been assigned values. Hence, the total number of possible injective solutions for the induced system of equations and non-equations is at least

$$\prod_{i=1}^{k}\left(2^n - \sigma_{i-1}w_i - \sum_{j=1}^{i-1}\tilde{w}_{ij} - cw_i\right).$$

Now, there could be the vertices which do not belong to the set $G^{=}(\mathcal{V})$. Let, there are k' such vertices. Fix such a vertex Y_{σ_k+i} and let us assume that w'_{σ_k+i} blue dashed edges are incident on Y_{σ_k+i}. If y_{σ_k+i} is a valid solution to the variable Y_{σ_k+i}, then we must have the following:

- y_{σ_k+i} should be distinct from previous σ_k assigned values.
- y_{σ_k+i} should be distinct from $(i-1)$ assigned values to the variables of the set $\mathcal{V} \setminus G^{=}(\mathcal{V})$.
- y_{σ_k+i} should not take w'_{σ_k+i} values such that it violates the non-equality conditions of w'_{σ_k+i} blue dashed edges.

Therefore, the number of valid choices of y_{σ_k+i} is at least $(2^n - \sigma_k - i + 1 - w'_{\sigma_k+i})$. Summarizing above, the total number of possible injective solutions for the induced system of equations and non-equations is at least

$$\prod_{i=1}^{k}\left(2^n - \sigma_{i-1}w_i - \sum_{j=1}^{i-1}\tilde{w}_{ij} - cw_i\right) \cdot \prod_{i\in[k']}(2^n - \sigma_k - i + 1 - w'_{\sigma_k+i})$$

which proves the result. □

Finalizing the proof of Theorem 3. From Lemma 5, the number of injective solutions to the system of equations $\mathcal{E}_G$ is at least $\prod_{i=1}^{k}(2^n - \sigma_{i-1}w_i - \tilde{w}_{i1} - \ldots - \tilde{w}_{i,i-1} - cw_i) \cdot \prod_{i\in[k']}(2^n - \sigma_k - i + 1 - w'_{\sigma_k+i})$, where w_i is the size of the i-th component $\mathcal{C}_i, \sigma_{i-1} = (w_1 + \ldots + w_{i-1}), k'$ is the number of vertices in $G \setminus G^{=}(\mathcal{V})$ and w'_{σ_k+i} is the number of blue dashed edges incident on the vertex Y_{σ_k+i}. Similar to the proof of Lemma 4, we derive the expression of Theorem 3 by doing the following algebra. For the notational simplicity, we denote $(\tilde{w}_{i1} + \ldots + \tilde{w}_{i,i-1})$ as p_i.

$$\begin{aligned} h_\alpha \frac{2^{nq_m}}{(2^n)_\alpha} &\geq \frac{2^{nq_m}}{(2^n)_\alpha}\prod_{i=1}^{k}(2^n - \sigma_{i-1}w_i - p_i - cw_i)\prod_{i=1}^{k'}(2^n - \sigma_k - i + 1 - w'_{\sigma_k+i}) \\ &= \underbrace{\prod_{i=1}^{k}\frac{(2^n - \sigma_{i-1}w_i - p_i - cw_i)2^{n(w_i-1)}}{(2^n - \sigma_{i-1})_{w_i}}}_{\mathsf{D.1}}\underbrace{\prod_{i=1}^{k'}\frac{(2^n - \sigma_k - i + 1 - w'_{\sigma_k+i})}{(2^n - \sigma_k - i + 1)}}_{\mathsf{D.2}}, \end{aligned}$$

where q_m is the number of edges of the subgraph $G^{=}$ and α is the number of vertices of G. In the following, we compute D.1 and D.2.

Compute $\mathsf{D.1}$. By doing the similar algebra as in the proof of Lemma 4, we have

$$\begin{aligned}
\mathsf{D.1} &\geq \prod_{i=1}^{k}\left(1 - \frac{A_i}{2^{2n} - 2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) + A_i} - \frac{2^n(p_i + cw_i)}{2^{2n} - 2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) + A_i}\right) \\
&\overset{(4)}{\geq} \prod_{i=1}^{k}\left(1 - \frac{2A_i}{2^{2n}} - \frac{2(p_i + cw_i)}{2^n}\right) \overset{(5)}{\geq} \left(1 - \sum_{i=1}^{k}\frac{6\sigma_{i-1}^2\binom{w_i}{2}}{2^{2n}} - \sum_{i=1}^{k}\frac{2(p_i + cw_i)}{2^n}\right) \\
&\overset{(6)}{\geq} \left(1 - \sum_{i=1}^{k}\frac{6\sigma_{i-1}^2\binom{w_i}{2}}{2^{2n}} - \frac{2q'_v}{2^n} - \frac{2c\alpha}{2^n}\right),
\end{aligned}$$

where (4) follows from the fact that $2^n(\sigma_{i-1}w_i + \binom{w_i}{2}) - A_i \leq 2^{2n}/2$, which holds to true when $\sigma_k w_{\max} \leq 2^n/4$, (5) holds to true due to the fact that $A_i \leq 3\sigma_{i-1}^2\binom{w_i}{2}$ and (6) holds to true as we denote $(p_1 + \ldots + p_k) = q'_v$, the total number of blue dashed edges across the components of $G^{=}$ and $w_1 + \ldots + w_k \leq \alpha$.

Compute $\mathsf{D.2}$. For computing $\mathsf{D.2}$, we have

$$\begin{aligned}
\mathsf{D.2} &= \prod_{i=1}^{k'}\frac{(2^n - \sigma_k - i + 1 - w'_{\sigma_k+i})}{(2^n - \sigma_k - i + 1)} \geq \prod_{i=1}^{k'}\left(1 - \frac{w'_{\sigma_k+i}}{(2^n - \sigma_k - i + 1)}\right) \\
&\overset{(7)}{\geq} \left(1 - \sum_{i=1}^{k'}\frac{2w'_{\sigma_k+i}}{2^n}\right) \overset{(8)}{\geq} \left(1 - \frac{2q''_v}{2^n}\right),
\end{aligned}$$

where (7) follows due to the fact that $(\sigma_k + i - 1) \leq 2^n/2$ and (8) follows as we denote $(w'_{\sigma_k+1} + \ldots + w'_{\sigma_k+k'}) = q''_v$, the total number of blue dashed edges incident on the vertices outside of the set $G^{=}(\mathcal{V})$.

Combining $\mathsf{D.1}$ and $\mathsf{D.2}$. Finally, by combining the expression of $\mathsf{D.1}$ and $\mathsf{D.2}$, we have

$$h_\alpha \frac{2^{nq_m}}{(2^n)_\alpha} \geq \left(1 - \sum_{i=1}^{k}\frac{6\sigma_{i-1}^2\binom{w_i}{2}}{2^{2n}} - \frac{2(q_v + c\alpha)}{2^n}\right),$$

where $q_v = q'_v + q''_v$, the total number of non-equation edges in G. □

C Proof of Claim 1

We recall claim 1 which says that if $\tau' = (\tau_m, \tau_v, K_h)$ is a good transcript then the induced graph $G_{\tau'}$ is valid. To prove that $G_{\tau'}$ is valid, we need to show (i) $G^{=}_{\tau'}$ is acyclic, (ii) $G_{\tau'}$ satisfies NPL condition and (iii) $G_{\tau'}$ satisfies NCL condition. For doing this, we inherit the notations introduced while analysing the probability of good transcripts in the proof of Theorem 1. $\widehat{N}_i$ denotes $0\|N_i$, $\widehat{X}_i$ denotes $1\|X_i$, $\widehat{N'}_a$ denotes $0\|N'_a$ and $\widehat{X'}_a$ denotes $1\|X'_a$ where $X_i = N_i \oplus \mathsf{H}_{K_h}(M_i)$.

(1) <u>$G^{=}_{\tau'}$ **is acylic.**</u> For the sake of contradiction, let us assume there is a cycle C in the graph $G^{=}_{\tau}$. If $|C| = 2$, then there must exist two authentication equations

$$\Pi(\widehat{N}_i) \oplus \Pi(\widehat{X}_i) = T_i, \qquad \Pi(\widehat{N}_j) \oplus \Pi(\widehat{X}_j) = T_j$$

in $\mathcal{E}_m$ with $N_i = N_j$ and $X_i = X_j$. But this event is nothing but the bad event (B2) in Definition 2. As the transcript τ' is good, this event cannot hold and therefore, there cannot be any cycle of length 2 in $G^{=}_{\tau'}$. A careful observation reveals that there cannot be any cycle of length 3 in the graph. Moreover, if there is any cycle of length at least 4 in $G^{=}_{\tau'}$, there must exist three authentication equations

$$\Pi(\widehat{N}_i) \oplus \Pi(\widehat{X}_i) = T_i, \qquad \Pi(\widehat{N}_j) \oplus \Pi(\widehat{X}_j) = T_j, \qquad \Pi(\widehat{N}_k) \oplus \Pi(\widehat{X}_k) = T_k$$

in $\mathcal{E}_m$ with $N_i = N_j$ and $X_j = X_k$. But this event is simply the bad event (B2) as Definition 2. As the transcript τ' is good, this event cannot hold and therefore there cannot be any cycle in $G^{=}_{\tau'}$ with length at least 4. Summarizing above, graph $G^{=}_{\tau'}$ is acyclic.

(2) <u>$G_{\tau'}$ **satisfies NPL.**</u> First of all, label of each edge of the graph is non-zero as τ' is good. Now, we consider any path P of length 2 in $G^{=}_{\tau}$. Let the label of the edges of the path be T_i and T_j. This implies that there must be two authentication equations

$$\Pi(\widehat{N}_i) \oplus \Pi(\widehat{X}_i) = T_i, \qquad \Pi(\widehat{N}_j) \oplus \Pi(\widehat{X}_j) = T_j$$

in $\mathcal{E}_m$ with $N_i = N_j$ or $X_i = X_j$. Now, if $T_i = T_j$, then this would create a cycle of length 2 in $G^{=}_{\tau'}$ which is impossible as we have proved that $G^{=}_{\tau'}$ is acyclic. Therefore, there cannot be any path of length 2 in $G^{=}_{\tau'}$ such that the label of the path becomes zero. Moreover, one cannot have any path of length at least 3 in $G^{=}_{\tau'}$; otherwise bad condition (B2) would have been satisfied. Therefore, $G_{\tau'}$ satisfies **Non-zero path label** condition.

(3) <u>$G_{\tau'}$ **satisfies NCL.**</u> Consider first, a cycle of length 2 where one edge is a blue dotted edge. Then there must be one authentication equation and one verification non-equation

$$\Pi(\widehat{N}_i) \oplus \Pi(\widehat{X}_i) = T_i, \qquad \Pi(\widehat{N'}_a) \oplus \Pi(\widehat{X'}_a) \neq T'_a$$

such that $N_i = N'_a, X_i = X'_a$ and $T_i = T'_a$. But this implies that the event actually satisfies the bad condition (B4) in Definition 2. As the transcript τ' is good, this event cannot hold and therefore, there cannot be any cycle of length 2 with one blue dotted edge. Moreover, as we have argued before, there cannot be any cycle of length 3 with exactly one non-equation edge. Now, for the existence of a cycle with length at least 4 that contains exactly one non-equation edge, there must be a path with at length least 3 in $G^{=}_{\tau'}$. But clearly, this is impossible. In summary, $G_{\tau'}$ satisfies **Non-zero cycle label** condition. □

D Details Proof of Theorem 2

We start by recalling the definition of bad transcripts and the lemma for bounding the probability of bad transcripts in the ideal world. An attainable transcript $\tau' = (\tau_e, \tau_d, K_h, \widetilde{\mathbf{T}})$ is said to be **bad** if any one of the following conditions occurs:

- B.1 : $\exists\ i \in [q_e]$ and $j \in [l_i]$ such that $S_i[j] = K_h$.
- B.2 : $\exists\ i \in [q_e]$ and $j \in [l_i]$ such that $S_i[j] = \mathbf{0}$.
- B.3 : $\exists\ i \in [q_e]$ and $j, j' \in [l_i]$ such that $S_i[j] = S_i[j']$.
- B.4 : $\exists\ i \in [q_e]$ such that $\widetilde{T}_i = \mathbf{0}$.
- B.5 : $\exists i \neq j, j \neq k$ such that $\widehat{N}_i = \widehat{N}_j$ and $\widehat{X}_j = \widehat{X}_k$.
- B.6 : $\{i_1, \ldots, i_{\xi+1}\} \subseteq [q_e]$ such that $\widehat{X}_{i_1} = \widehat{X}_{i_2} = \ldots = \widehat{X}_{i_{\xi+1}}$ for some parameter $\xi \geq \mu$.
- B.7 $\exists\ a \in [q_d]$, $\exists\ i \in [q_e]$ such that $\widehat{N}_i = \widehat{N'}_a$, $\widehat{X}_i = \widehat{X'}_a$ and $\widetilde{T}_i = T'_a$.

Θ_{bad} (resp. Θ_{good}) denotes the set of bad (resp. good) transcripts. Moreover, X_{re} and X_{id} denotes the probability distribution of realizing an extended transcript τ' in the real and the ideal world respectively. We once again recall Lemma 3 to bound the probability of bad transcripts in the ideal world

Lemma 4. *Let X_{id} and Θ_{bad} be defined as above. Then*

$$\Pr[X_{\text{id}} \in \Theta_{\text{bad}}] \leq \epsilon_{\text{bad}} = \frac{2\sigma}{2^n} + \frac{q_e \ell^2}{2^n} + \frac{q_e}{2^n} + \frac{q_e^2 \ell}{\xi 2^n} + \frac{(2q_e + q_d)2\ell\mu}{2^n} + \frac{2q_d \ell}{2^n}.$$

Proof. By the union bound,

$$\Pr[X_{\text{id}} \in \Theta_{\text{bad}}] \leq \sum_{i=1}^{7} \Pr[\mathsf{B.i}]. \tag{39}$$

In the following, we only bound $\Pr[\mathsf{B.1}], \Pr[\mathsf{B.2}]$ and $\Pr[\mathsf{B.3}]$ as the bound for the remaining events can be found in the proof of Lemma 2. Clearly,

$$\Pr[\mathsf{B.1}] \leq \frac{\sigma}{2^n}. \tag{40}$$

Bounding B.2. Event B.2 occurs if there exists a zero keystream block in any query. For a fixed query and a block, the probability of this event holds is exactly 2^{-n}. When the i-th query is not faulty, then the probability of any block takes the output $\mathbf{0}$ is exactly 2^{-n}. If the i-th query is faulty, then we have the following two cases:

- **Case (i):** When the j-th block is sampled in executing the i-th query, then the probability is 2^{-n}.
- **Case (ii):** When the j-th block is not sampled in executing the i-th query. That implies there must be some previous encryption query for which the j-th block is freshly sampled and hence the probability is 2^{-n}.

By summing over all choices of i and j, we bound the probablity to $\sigma/2^n$. Therefore, we have

$$\Pr[\mathsf{B.2}] \leq \frac{\sigma}{2^n}. \tag{41}$$

Bounding B3. Event $\mathsf{B3}$ occurs if there is a collision in two different keystream blocks in an encryption query. For a fixed query and two fixed distinct blocks, the probability of this event holds is exactly 2^{-n}. When the i-th query is not faulty, then the probability of such collision is exactly 2^{-n}. If the i-th query is faulty, then we have the following two cases:

- **Case (i):** When either of the blocks is sampled in executing the i-th query, then the probability is 2^{-n}.
- **Case (ii):** When none of the two blocks are sampled in executing the i-th query, it means that there must be some previous encryption query for which either of the blocks was freshly sampled, and hence the probability is 2^{-n}.

By summing over all choices of i, j and j', we bound the probability to at most $q_e\ell^2/2^n$. Therefore, we have

$$\Pr[\mathsf{B.3}] \leq \frac{q_e\ell^2}{2^n}. \tag{42}$$

From Lemma 2, we obtain the bound of $\Pr[\mathsf{B.4}] + \Pr[\mathsf{B.5}] + \Pr[\mathsf{B.6}] + \Pr[\mathsf{B.7}]$. Therefore, from Eqn. (39), Eqn. (40), Eqn. (41), Eqn. (42) and from Lemma 2, the result follows with $\epsilon \leq 2\ell/2^n$; almost xor universal probability of the truncated PolyHash. □

D.1 Analysis of Good Transcripts

In this section, we show that for a good transcript $\tau' = (\tau, K_h, \widetilde{\mathbf{T}})$, realizing τ' is almost as likely in the real world as in the ideal world. Formally, we prove the following lemma.

Lemma 6. *Let $\tau' = (\tau_e, \tau_d, K_h, \widetilde{\mathbf{T}})$ be a good transcript. Then*

$$\frac{\Pr[X_{\mathrm{re}} = \tau']}{\Pr[X_{\mathrm{id}} = \tau']} \geq \left(1 - \sum_{i=1}^{k} \frac{24\sigma^2\binom{W_i}{2}}{2^{2n}} - \frac{2q_d}{2^\rho} - \frac{2\sigma}{2^n}\right),$$

where σ is the number of message blocks queried and ρ is the size of the tag.

Proof. Let us consider a good transcript $\tau' = (\tau_e, \tau_d, K_h, \widetilde{\mathbf{T}})$. Since in the ideal world the encryption oracle is perfectly random and the decryption oracle always rejects, one simply has

$$\Pr[X_{\mathrm{id}} = \tau'] = \frac{1}{2^n} \cdot \prod_{t=1}^{r} \frac{1}{2^{nl_t}} \cdot \frac{1}{2^{nq_e}} \tag{43}$$

where r is the number of groups of nonces and l_t be the updated number of generated keystream blocks for group t. Now, we lower bound the probability of getting τ' in the real world. A permutation $\mathsf{\Pi}$ is *compatible with* τ_e if (A) happens

$$(\mathrm{A}) = \begin{cases} \forall i \in [q_e], j \in [l_i], \mathsf{\Pi}(\widehat{N}_i) \oplus \mathsf{\Pi}(0 \| N_i \| \langle j \rangle) = S_i[j] \\ \forall i \in [q_e] \mathsf{\Pi}(\widehat{N}_i) \oplus \mathsf{\Pi}(\widehat{X}_i) = T_i \end{cases}$$

and *compatible with* τ_d if (B) happens

$$(\mathrm{B}) = \ \forall a \in [q_d], \mathsf{\Pi}(\widehat{N'}_a) \oplus \mathsf{\Pi}(\widehat{X'}_a) \neq T'_a \| \beta,$$

where recall that $\widehat{N}_i = 0 \| N_i$, $\widehat{X}_i = 1 \| X_i$, $\widehat{N'}_a = 0 \| N'_a$ and $\widehat{X'}_a = 1 \| X'_a$ and $\beta \in \{0,1\}^{n-\rho}$. Moreover, $\langle j \rangle$ denotes the $n/4 - 1$ bit binary representation of non-zero integer j. $\mathsf{\Pi}$ is compatible with τ' if it is compatible with τ_e and τ_d. Let $\mathsf{Comp}(\tau)$ denote the set of all permutations that are compatible with τ. Therefore,

$$\begin{aligned} \mathsf{p}_{\mathrm{re}}(\tau) \overset{\Delta}{=} \Pr[X_{\mathrm{re}} = \tau'] &= \frac{1}{|\mathcal{K}_h|} \cdot \Pr[\mathsf{\Pi} \leftarrow_{\$} \mathsf{Perm} : \mathsf{\Pi} \in \mathsf{Comp}(\tau)] \\ &= 2^{-n} \cdot \underbrace{\Pr[(\mathrm{A}), (\mathrm{B}) \text{ holds}]}_{\mathsf{P}_{ed}}, \end{aligned} \tag{44}$$

where the third equality occurs due to the randomness of the hash key $\mathsf{E}_K(\mathbf{0})$. Now. if we model the system of equations and non-equations into the graph theoretic setting, one can see that graph $G_{\tau'}$ generated out of the transcript τ' is valid (i.e., it satisfies NC, NPL and NCL conditions). By assuming $\xi \leq 2^n/8\sigma\ell$ (follows from Theorem 3), we apply Theorem 3 with $c = 1$ and assuming $\alpha \leq \sigma$, to obtain,

$$\mathsf{P}_{ed} \geq \frac{1}{2^{nq_e}} \prod_{t=1}^{r} \frac{1}{2^{nl_t}} \cdot \left(1 - \sum_{i=1}^{k} \frac{6\sigma'^2_{i-1}\binom{W'_i}{2}}{2^{2n}} - \frac{2q_d}{2^\rho} - \frac{2\sigma}{2^n}\right), \tag{45}$$

where k is the number of components of $G_{\tau'}$, W'_i denotes the size of the i-th component and $\sigma'_i = W'_1 + \ldots W'_i$. Note that there are $2^{n-\rho} q_d$ non-equations as the adversary forges with ρ bit tags T'_a and there are $2^{n-\rho}$ $\widetilde{T}$'s whose first ρ bits matches with T'_a. Moreover, we consider $c = 1$ due to the fact that we choose elements from the set $\{0,1\}^n$ excluding the hash key. Finally, the result follows from Eqn. (43), Eqn. (44), Eqn. (45) and the inequality $\sigma'_{i-1} \leq \sigma'_k = (W'_1 + \ldots + W'_k) \leq 2\sigma$. □

Finalizing the proof. From Lemma 6, we define $\phi(\tau')$ as follows:

$$\phi(\tau') \overset{\Delta}{=} \sum_{i=1}^{k} \frac{24\sigma^2\binom{W'_i}{2}}{2^{2n}} + \frac{2q_d}{2^\rho} + \frac{2\sigma}{2^n}.$$

Now, we calculate the expectation of $\phi(\tau')$ as follows:

$$\mathbf{E}[\phi(X_{\mathrm{id}})] = \left(\frac{2q_d}{2^\rho} + \frac{2\sigma}{2^n} + \frac{24\sigma^2}{2^{2n}}\mathbf{E}\left[\sum_{i=1}^{k}\binom{W_i'}{2}\right]\right). \tag{46}$$

Moreover, one can easily see the following inequality

$$\binom{W_i'}{2} \leq \binom{W_i}{2}\binom{2\ell}{2}, \tag{47}$$

where ℓ is the maximum number of message blocks and W_i is as defined in the proof of Theorem 1. Therefore, from Eqn. (24) and Eqn. (47), we have

$$\mathbf{E}\left[\sum_{i=1}^{k}\binom{W_i'}{2}\right] \leq 2\ell^2\mathbf{E}\left[\sum_{i=1}^{k}\binom{W_i}{2}\right] \leq 2\ell^2\mathbf{E}\left[\sum_{i=1}^{k}\binom{\tilde{W}_i}{2}\right] + 4q_e\ell^2 \tag{48}$$

Moreover, from Eqn. (25) we have

$$\mathbf{E}\left[\sum_{i=1}^{k}\binom{\tilde{W}_i}{2}\right] \leq \sum_{i\neq j}^{q_e}\mathbf{E}[I_{ij}] + \sum_{i\neq j}^{\mu}\mathbf{E}[\tilde{I}_{ij}] \leq q_e^2\ell/2^n + \mu^2/2 \tag{49}$$

where the almost xor universal probability of the truncated PolyHash is at most $2\ell/2^n$. Finally, from Eqn. (46), Eqn. (48) and Eqn. (49) we have

$$\mathbf{E}[\phi(X_{\mathrm{id}})] \leq \left(\frac{2q_d}{2^\rho} + \frac{2\sigma}{2^n} + \frac{48\sigma^4\ell}{2^{3n}} + \left(\frac{5\sigma\ell\mu}{2^n}\right)^2 + \frac{96\sigma^3\ell}{2^{2n}}\right), \tag{50}$$

where we assume that $\ell q_e \approx \sigma$, the total number of message blocks queried. Finally, by applying Eqn. (2), Lemma 3, Eqn. (50), $\xi = 2^n/8\sigma\ell$ and $\sigma \leq 2^n/48$, the proof follows. □