EWCDM: An Efficient, Beyond-Birthday Secure, Nonce-Misuse Resistant MAC*

Benoît Cogliati and Yannick Seurin

University of Versailles, France
benoitcogliati@hotmail.fr

ANSSI, Paris, France
yannick.seurin@m4x.org

May 27, 2016

Abstract. We propose a nonce-based MAC construction called EWCDM (*Encrypted Wegman-Carter with Davies-Meyer*), based on an almost xor-universal hash function and a block cipher, with the following properties: (i) it is simple and efficient, requiring only two calls to the block cipher, one of which can be carried out in parallel to the hash function computation; (ii) it is provably secure beyond the birthday bound when nonces are not reused; (iii) it provably retains security up to the birthday bound in case of nonce misuse. Our construction is a simple modification of the Encrypted Wegman-Carter construction, which is known to achieve only (i) and (iii) when based on a block cipher. Underlying our new construction is a new PRP-to-PRF conversion method coined *Encrypted Davies-Meyer*, which turns a pair of secret random permutations into a function which is provably indistinguishable from a perfectly random function up to at least $2^{2n/3}$ queries, where n is the bit-length of the domain of the permutations.

Keywords: Wegman-Carter MAC, Davies-Meyer construction, nonce-misuse resistance, beyond-birthday-bound security

1 Introduction

Wegman-Carter MACs. A *Message Authentication Code* (MAC) is a fundamental symmetric-key primitive that allows a sender to authenticate messages by computing tags that can be verified by the receiver (the sender and the receiver sharing a common secret key). Many MACs are based on some underlying cryptographic primitive such as a block cipher (e.g., CBC-MAC [BKR00]) or a hash function (e.g., HMAC [BCK96]). A different approach, pioneered by Wegman and Carter [WC81] (building on earlier work by Gilbert, MacWilliams, and Sloane [GMS74]), first treats the message M with an almost xor-universal

(AXU) hash function[1] H (i.e., a fast, *combinatorial* primitive rather than a slow, *cryptographic* one) and masks the result with a one-time pad, resulting in *information-theoretically secure* authentication. Since sharing a one-time pad for each message to authenticate is not very practical, one can instead use a pseudo-random function F, as first proposed by Brassard [Bra82], allowing the sender and the receiver to share a short secret K rather than a long list of one-time pads. The mask for each new message is then generated pseudorandomly by applying F_K to a *nonce* N, a value used at most once. This reintroduces a cryptographic primitive (and hence a computational assumption), but only for treating a small nonce rather than a potentially long message. The resulting nonce-based MAC, that we simply call the *Wegman-Carter* (WC) construction, is

$$\mathsf{WC}[F,H]_{K,K_h}(N,M) = F_K(N) \oplus H_{K_h}(M),$$

where K is the key for the pseudorandom function F, K_h is the key for the AXU hash function H, N is the nonce, and M is the message.[2]

The WC construction enjoys a very strong provable security bound when nonces are never reused. Assuming that F is perfect (i.e., F_K is a uniformly random function), any adversary seeing at most q_m honestly generated tags and making at most q_v verification queries (i.e., forgery attempts) succeeds with probability at most εq_v, where ε is the maximal differential probability of H, namely

$$\varepsilon = \max_{X \neq X', Y} \Pr\left[H_{K_h}(X) \oplus H_{K_h}(X') = Y\right],$$

the probabilities being taken over the random draw of the hashing key K_h. When F is not perfect, there is an additional term accounting for its insecurity as a PRF (more precisely, this corresponds to the best advantage an adversary can achieve in distinguishing F_K from a uniformly random function within $q_m + q_v$ queries).

Many AXU hash functions have been proposed for instantiating this construction, most of them based on polynomial hashing [Kra94, Rog95, Sho96, HK97, BHK+99, Ber00, KR00, KVW04, MV04, Ber05c]. See [Ber07] for more references and a comprehensive survey of polynomial hashing. Universal hash functions can also be constructed from a block cipher (e.g. by using the CBC mode with prefix-free encoding [BR05, BPR05]), but in that case the provable maximal differential probability depends on the PRP-security of the block cipher (hence, this yields "computational" rather than "statistical" universal hash functions).

Nonce-Misuse Resistance. Despite the advantages just mentioned (efficiency and excellent security bound), the WC construction has one major shortcoming:

[1] An AXU hash function is a keyed function with the property that for any two distinct inputs, the probability over the draw of a random key that the outputs have a specific difference is small.

[2] Here and in all the following, we assume to fix ideas that the outputs of the PRF and the hash function are n-bit strings and the group operation is bitwise xor; this can be easily adapted to any other abelian group.

it is very vulnerable to *nonce-misuse*. If a nonce is repeated even a single time, consequences can be catastrophic [Jou06, HP08]. For example, in the case of polynomial universal hashing, this can lead to a complete recovery of the hashing key, which allows universal forgeries. To remedy this nonce-misuse problem, the simplest option, which has been known for long, is to apply the PRF to the output of the hash function. For instance, if the PRF takes $2n$-bit inputs, one can define the tag as $F_K(N \| H_{K_h}(M))$; this construction was analyzed by Black *et al.* [BHK+99, BC09]. If F takes only n-bit inputs, one can instead apply the PRF with an independent key to the output of the WC construction, thereby defining the tag as

$$F_{K'}\big(F_K(N) \oplus H_{K_h}(M)\big). \tag{1}$$

If one gets rid of the nonce, simply defining the tag as $F_K(H_{K_h}(M))$, one obtains a stateless MAC but the security bound includes an extra "birthday-type" term εq_m^2.

Beyond-Birthday-Bound Security. There is another obstacle which can prevent concrete implementations from enjoying the strong security bound promised by the WC construction: pseudorandom functions are not always readily available, and it is common to use a pseudorandom *permutation* instead, or in other words to replace F with a block cipher E. However, as first pointed out by Shoup [Sho96], this causes the proven security bound to drop to the so-called birthday bound. Indeed, a random permutation can be distinguished from a random function within q queries with advantage roughly $q^2/2^n$. For resource-constrained environments, where lightweight cryptographic primitives based on block ciphers with 64-bit blocks are likely to be implemented, this means that security insurance is lost after 2^{32} queries, which is often unacceptable, especially when refreshing keys regularly is excluded.

A first solution to overcome the birthday bound while using only a block cipher is to use a *randomized* construction. However, existing schemes either require very strong properties from the block cipher such as the ideal cipher model [JJV02] or resistance to related-key attacks [JL04], or require a relatively large amount of randomness (at least $3n$ bits for the MACRX construction of [BGK99]). The beyond-birthday-bound secure construction named MAC-R2 of Minematsu [Min10] uses a random n-bit IV per message and bears resemblance to the construction proposed in this paper, but it requires four calls to the underlying block cipher. (Jumping ahead, our new construction requires only two calls.) Moreover, reliable randomness might not always be available in some environments, and it might sometimes be easier to maintain a state.

Another option is to implement F_K in construction (1) from a block cipher E using a so-called *PRP-to-PRF conversion method* [BKR98, HWKS98] with beyond-birthday-bound security. (On the other hand, it is easy to see that the outer PRF $F_{K'}$ can be directly implemented by a block cipher without security loss.) Perhaps the simplest such method is the "xor" construction $E_{K_1}(N) \oplus E_{K_2}(N)$, or its close single-key variant $E_K(N\|0) \oplus E_K(N\|1)$, which have been analyzed in a number of papers [BI99, Luc00, Pat08a, Pat13, CLP14].

However, all known methods require at least two block cipher calls; taking into account the outer encryption layer, this amounts to three block cipher calls for the whole construction. Is it possible to do better?

Our Contribution. We propose a new nonce-based MAC based on a AXU hash function and a block cipher with the following properties:

(i) it is simple and efficient, requiring only two calls to the underlying block cipher, one of which can be carried out in parallel to the hash function computation;
(ii) it provably provides security *beyond the birthday bound* when nonces are never reused;
(iii) it provably retains security up to the birthday bound in case of nonce misuse.

Property (ii) ensures that the scheme is highly secure in the nominal use case where nonces are never repeated, while property (iii) acts as a "safety net" if anything goes wrong with nonces.

Our starting point is what we call the Encrypted Wegman-Carter construction, which is simply construction (1) where the outer PRF layer is replaced by a block cipher, viz.

$$E_{K'}\big(F_K(N) \oplus H_{K_h}(M)\big). \tag{2}$$

As already briefly explained, this construction enjoys the same security bound as the (unencrypted) WC construction when nonces are never repeated, and is moreover nonce-misuse resistant up to the birthday bound. Replacing F_K by a simple block cipher call causes the security bound to drop to the birthday bound even when nonces are not repeated, while using a PRP-to-PRF conversion method with security beyond the birthday bound results in at least three block cipher calls in total for the resulting construction.

Our main observation is that one can overcome the birthday bound in the nonce-respecting scenario by instantiating F_K using "only" the Davies-Meyer (DM) construction. The DM construction is the easiest way to turn a block cipher into a keyed function.[3] Given a block cipher $E : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$, the DM construction based on E is simply

$$\mathsf{DM}[E]_K(N) = E_K(N) \oplus N.$$

Note that this PRF construction is *not* secure beyond the birthday bound: given black-box access to a function $f : \{0,1\}^n \to \{0,1\}^n$, a distinguisher can simply query $f(N_i)$ for roughly $2^{n/2}$ distinct values N_i and look for collisions in values $f(N_i) \oplus N_i$. When f is a uniformly random function this will happen with good probability, whereas when $f = \mathsf{DM}[E]_K$ this cannot happen. However, this attack is not possible anymore if one encrypts the output of the DM construction.

[3] Traditionally, the DM construction is rather seen as a way to turn a block cipher into an (unkeyed) compression function.

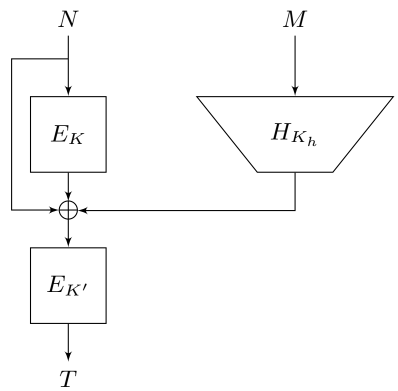

Fig. 1. The “Encrypted Wegman-Carter with Davies-Meyer” construction.

Using the DM construction to instantiate F_K in construction (2) results in a MAC construction based only on E and H, which we call *Encrypted Wegman-Carter with Davies-Meyer* (EWCDM) construction, depicted on Fig. 1 and defined as

$$E_{K'}\big(E_K(N) \oplus N \oplus H_{K_h}(M)\big). \tag{3}$$

Our main result is that the EWCDM construction is secure up to roughly $2^{2n/3}$ MAC queries and 2^n verification queries against nonce-respecting adversaries (while against nonce-misusing adversaries it still enjoys birthday-bound security). We stress that this does not hold for the (unencrypted) Wegman-Carter construction with Davies-Meyer: if tags are computed as

$$T = E_K(N) \oplus N \oplus H_{K_h}(M),$$

then the resulting MAC scheme is only provably secure up to the birthday bound against nonce-respecting adversaries.[4] Hence, the outer encryption layer $E_{K'}$ turns out to be *twice* useful: for providing nonce-misuse resistance on one hand, and for cheaply enhancing security against nonce-respecting adversaries beyond the birthday bound on the other hand.

We believe that our new construction would be an elementary and easy-to-implement way to enhance the security of widely deployed authentication or authenticated encryption schemes such as Poly1305-AES [Ber05c] or GCM [MV04] (in particular, note that this can be done in a black-box way on top of an existing implementation of those schemes). The main cost would be some additional latency due to the extra block cipher call, but depending on the context this might be tolerable.

[4] Indeed, the outputs of this construction can be distinguished from random simply by querying the MAC oracle for tags T_i with the same message and roughly $2^{n/2}$ distinct nonces N_i, and looking for collisions in $T_i \oplus N_i$.

Proof Technique. At the heart of construction (3) is a novel PRP-to-PRF conversion method: namely, if we make abstraction for a moment of the hash of the message M, and if we simply denote P and P' in place of E_K and $E_{K'}$, we obtain a function of the nonce defined as

$$F(N) = P'(P(N) \oplus N).$$

For obvious reasons, we call this the *Encrypted Davies-Meyer* (EDM) construction. The main part of the proof consists in proving that this is a secure PRF up to $2^{2n/3}$ adversarial queries. (We prove this as a standalone result in Appendix A; this constitutes a good warm-up for the reader before the more complicated security proof of the EWCDM construction in Section 4.) However, since the hash of the message is "intermingled" within the EDM construction, it does not seem possible to first prove that the outputs of the MAC oracle are indistinguishable from random, and then handle verification queries (as is usually done for proving the security of the standard Wegman-Carter construction; see Theorem 1 in Section 3.1). Note that one cannot hope either to prove security beyond the birthday bound by a sequence of games that would start by replacing the DM construction $E_K(N) \oplus N$ by a uniformly random function.

Hence, it seems that any proof aiming at security beyond the birthday bound must handle MAC queries *and* verification queries both at the same time. For this, we employ the H-coefficients technique, which has been introduced by Patarin [Pat90, Pat91, Pat08b] and which recently regained attention since Chen and Steinberger used it to analyze the iterated Even-Mansour cipher [CS14]. This technique gives a kind of "systematic" way to upper bound the statistical distance between the answers of two interactive systems and is typically used to prove (information-theoretic) pseudorandomness of constructions such as Feistel networks. To the best of our knowledge, this is the first time the H-coefficients technique is used for proving the security of a MAC (i.e., unpredictability rather than pseudorandomness).

More Related Work. This paper focuses on nonce-based (hence stateful) MACs, but there is also an important line of work aiming at constructing stateless and deterministic MACs secure beyond the birthday bound. However, existing constructions [Yas10, Yas11, DS11, ZWSW12] are far more complex than the one presented in this paper. We mainly mentioned works related to provable security; there is also a large number of papers (motivated by the analysis of the widely deployed GCM mode [MV04]) investigating attacks against polynomial hash-based MACs [Fer05, HP08, Saa12, PC15, ABBT15].

Open Problems. We prove the security of the EWCDM construction in the nonce-respecting scenario up to $2^{2n/3}$ MAC queries, but we conjecture that security actually holds up to close to 2^n queries (a similar conjecture holds for the Encrypted Davies-Meyer construction).

The EWCDM construction uses two distinct keys for the two calls to the block cipher; a natural question is whether security beyond the birthday bound also

Table 1. Proven security bounds (omitting constants and the term accounting for the PRP-security of the underlying block cipher) for the Wegman-Carter construction $\mathsf{WC}[E, H]$, the Encrypted Wegman-Carter construction $\mathsf{EWC}[E, H]$, and the new Encrypted Wegman-Carter with Davies-Meyer construction $\mathsf{EWCDM}[E, H]$.

	nonce-respecting	nonce-misusing
$\mathsf{WC}[E, H]$	$(q_m + q_v)^2/2^n + \varepsilon q_v$	—
$\mathsf{EWC}[E, H]$	$(q_m + q_v)^2/2^n + \varepsilon q_v$	$(q_m + q_v)^2/2^n + \varepsilon(q_m + q_v)^2$
$\mathsf{EWCDM}[E, H]$	$q_m^{3/2}/2^n + \varepsilon q_m + q_v/2^n + \varepsilon q_v$	$(q_m + q_v)^2/2^n + \varepsilon(q_m + q_v)^2$

holds when the same key is used. We believe this to be true, but likely cumbersome to prove. The corresponding question regarding the Encrypted Davies-Meyer construction is even more intriguing: How many queries are required to distinguish $P(x \oplus P(x))$ from a random function? It might well be that this construction is secure up to close to 2^n queries, which would yield the first optimally secure PRP-to-PRF conversion method which uses a single permutation (unlike $P_1(x) \oplus P_2(x)$) and does not shrink the domain (unlike $P(x\|0) \oplus P(x\|1)$).

Finally, it would be interesting to investigate how the security of EWCDM is affected by tag truncation. We believe that the only change to be made to the bound of Theorem 3 is to replace the term $6q_v/2^n$ by a term $O(q_v/2^\ell)$, where ℓ is the length of the truncated tag, but this remains to be proven.

Organization. We first establish the notation and recall standard security definitions in Section 2. In Section 3, we recall the previous security results on the Wegman-Carter and the Encrypted Wegman-Carter constructions, and describe our new EWCDM construction. We then prove the security of EWCDM in the nonce-respecting scenario in Section 4 and in the nonce-misusing scenario in Section 5. We also analyze the Encrypted Davies-Meyer PRP-to-PRF conversion method in Appendix A.

2 Preliminaries

Basic Notation. Given a non-empty set $\mathcal{X}$, we denote $X \leftarrow_\$ \mathcal{X}$ the draw of an element X from $\mathcal{X}$ uniformly at random. The set of all functions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted $\mathsf{Func}(\mathcal{X}, \mathcal{Y})$, and the set of all permutations of $\mathcal{X}$ is denoted $\mathsf{Perm}(\mathcal{X})$. The set of binary strings of length n is denoted $\{0,1\}^n$. The set of all functions from $\{0,1\}^n$ to $\{0,1\}^n$ is simply denoted $\mathsf{Func}(n)$, and the set of all permutations of $\{0,1\}^n$ is simply denoted $\mathsf{Perm}(n)$. For integers $1 \leq b \leq a$, we will write $(a)_b = a(a-1)\cdots(a-b+1)$ and $(a)_0 = 1$ by convention. Note that the probability that a random permutation $P \leftarrow_\$ \mathsf{Perm}(n)$ satisfies q equations $P(X_i) = Y_i$ for distinct X_i's and distinct Y_i's is exactly $1/(2^n)_q$.

PRFs and Block Ciphers. A keyed function with key space $\mathcal{K}$, domain $\mathcal{X}$, and range $\mathcal{Y}$ is a function $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$. We denote $F_K(X)$ for $F(K, X)$. A (q, t)-adversary against F is an algorithm A with oracle access to a function from $\mathcal{X}$ to $\mathcal{Y}$, making at most q oracle queries, running in time at most t, and outputting a single bit. The advantage of A in breaking the PRF-security of F is defined as

$$\mathbf{Adv}_F^{\mathrm{PRF}}(\mathsf{A}) = \left|\Pr\left[K \leftarrow_\$ \mathcal{K} : \mathsf{A}^{F_K} = 1\right] - \Pr\left[R \leftarrow_\$ \mathsf{Func}(\mathcal{X}, \mathcal{Y}) : \mathsf{A}^R = 1\right]\right|.$$

A block cipher with key space $\mathcal{K}$ and domain $\mathcal{X}$ is a mapping $E : \mathcal{K} \times \mathcal{X} \to \mathcal{X}$ such that for any key $K \in \mathcal{K}$, $X \mapsto E(K, X)$ is a permutation of $\mathcal{X}$. We denote $E_K(X)$ for $E(K, X)$. A (q, t)-adversary against E is an algorithm A with oracle access to a permutation of $\mathcal{X}$, making at most q oracle queries, running in time at most t, and outputting a single bit. The advantage of A in breaking the PRP-security of E is defined as

$$\mathbf{Adv}_E^{\mathrm{PRP}}(\mathsf{A}) = \left|\Pr\left[K \leftarrow_\$ \mathcal{K} : \mathsf{A}^{E_K} = 1\right] - \Pr\left[P \leftarrow_\$ \mathsf{Perm}(\mathcal{X}) : \mathsf{A}^P = 1\right]\right|.$$

Note that we do not need the strongest "two-sided" version of PRP-security (where the adversary also has access to a decryption oracle) since all constructions considered in this paper only use the forward (encryption) direction of the underlying block cipher.

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

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

$$\mathbf{Adv}_F^{\mathrm{MAC}}(\mathsf{A}) = \Pr\left[K \leftarrow_\$ \mathcal{K} : \mathsf{A}^{F_K, \mathsf{Ver}_K} \text{ forges}\right],$$

where the probability is also taken over the random coins of A, if any. The adversary is not allowed to ask a verification query (N, M, T) if a previous query (N, M) to F_K returned T. The adversary is said nonce-respecting *if it never repeats a nonce $N \in \mathcal{N}$ in its queries to the first oracle F_K.*

We say that an adversary is *nonce-misusing* if it does not abide to the rule of non-repeating nonces. The MAC-security of F in face of nonce-misusing adversaries is defined exactly as above, and can be rephrased as the standard (i.e., not nonce-based) MAC-security of a keyed function with domain $\mathcal{N} \times \mathcal{M}$.

AXU Hash Functions. We will need the following definition of an almost xor-universal (AXU) hash function.

Definition 2 (ε-AXU Hash Function). *Let $\mathcal{K}_h$, $\mathcal{X}$ and $\mathcal{Y}$ be three non-empty sets and $\varepsilon > 0$. A keyed function $H : \mathcal{K}_h \times \mathcal{X} \to \mathcal{Y}$ is said to be ε-AXU if for any distinct $X, X' \in \mathcal{X}$ and any $Y \in \mathcal{Y}$,*

$$\Pr[K_h \leftarrow_{\$} \mathcal{K}_h : H_{K_h}(X) \oplus H_{K_h}(X') = Y] \leq \varepsilon.$$

3 Wegman-Carter MAC Constructions

3.1 The Standard Wegman-Carter Construction

We recall the standard Wegman-Carter construction [WC81] of a nonce-based MAC from an ε-AXU hash function and a PRF. Let $\mathcal{K}$, $\mathcal{K}_h$, and $\mathcal{M}$ be non-empty sets. Let $F : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a keyed function and $H : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ε-AXU hash function. The Wegman-Carter construction based on F and H is the nonce-based keyed function with key space $\mathcal{K} \times \mathcal{K}_h$, nonce space $\{0,1\}^n$, message space $\mathcal{M}$, and range $\{0,1\}^n$ defined by

$$\mathsf{WC}[F,H]_{K,K_h}(N,M) = F_K(N) \oplus H_{K_h}(M).$$

We recall the classical security result for this construction [WC81] and sketch the proof for completeness. Here and in all the following, t_H is an upper bound on the time needed to compute $H_{K_h}(M)$ for any key $K_h \in \mathcal{K}_h$ and any message $M \in \mathcal{M}$.

Theorem 1. *Let F and H be as above. Then for any (q_m, q_v, t)-nonce-respecting adversary A against the MAC-security of $\mathsf{WC}[F,H]$, there exists a $(q_m + q_v, t')$-adversary A' against the PRF-security of F, where $t' = O(t + (q_m + q_v)t_H)$, such that*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{WC}[F,H]}(\mathsf{A}) \leq \mathbf{Adv}^{\mathrm{PRF}}_{F}(\mathsf{A}') + \varepsilon q_v.$$

Proof. Fix a (q_m, q_v, t)-nonce-respecting adversary A. Consider the WC construction where F_K is replaced by a uniformly random function R, and let δ be the advantage of A against this new construction. By a straightforward hybrid argument, there is an adversary A', making at most $q_m + q_v$ oracle queries, and running in time $O(t + (q_m + q_v)t_H)$, such that

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{WC}[F,H]}(\mathsf{A}) \leq \mathbf{Adv}^{\mathrm{PRF}}_{F}(\mathsf{A}') + \delta.$$

The answers $R(N) \oplus H_{K_h}(M)$ of the MAC oracle are now uniformly random and independent from K_h. Consider the i-th verification query (N', M', T') of the adversary. If N' never appeared in the MAC queries of the adversary, then T' is valid with probability 2^{-n}. If $N' = N$ for some previous MAC query (N, M) that returned T, then the verification query is valid iff

$$R(N') \oplus H_{K_h}(M') = T' \Leftrightarrow H_{K_h}(M) \oplus H_{K_h}(M') = T \oplus T',$$

which happens with probability at most ε by definition of an ε-AXU hash function. (If $M = M'$, then one must have $T \neq T'$ by definition of the security experiment, and the forgery cannot be valid.) Since for an ε-AXU hash function with range $\{0,1\}^n$ one has $\varepsilon \geq 2^{-n}$, in all cases the forgery is valid with probability at most ε. By a union bound over the q_v verification queries, one has $\delta \leq \varepsilon q_v$, which concludes the proof. □

Assume now that F is a family of *permutations* of $\{0,1\}^n$, or in other words, a block cipher, that we denote E. Then E can be distinguished from a random function with q queries and advantage roughly $q^2/2^n$ by simply looking for collisions in its outputs. In other words, by the PRP-PRF switching lemma [BR06], the best upper bound one can hope to prove for the PRF-advantage of adversary A' appearing in Theorem 1, assuming that E is a secure *PRP*, is

$$\mathbf{Adv}_E^{\mathrm{PRF}}(\mathsf{A}') \leq \mathbf{Adv}_E^{\mathrm{PRP}}(\mathsf{A}') + \frac{(q_m + q_v)^2}{2^{n+1}},$$

so that the security bound for the resulting construction $\mathsf{WC}[E, H]$ now has a birthday-type term. Bernstein [Ber05a, Ber05b] proved a better (but still of birthday-type) bound: as long as $q_m \leq 2^{n/2}$, the adversary can forge with probability at most $C\varepsilon q_v$, for some small constant C (in all practical cases, $C \leq 2$). Note that the distinguishing attack against E does not seem to translate into a forgery attack against the MAC scheme, and it might be possible to improve the security bound under additional assumptions on H and E.

3.2 Nonce-Misuse Resistance and the Encrypted Wegman-Carter Construction

In general, the standard Wegman-Carter construction of the previous section does not offer any security against nonce-misusing adversaries. Consider for example the case where H is a polynomial-based hash function. Then any adversary who gets two tags T and T' for two different messages M and M' generated with the same nonce knows that $H_{K_h}(M) \oplus H_{K_h}(M') \oplus T \oplus T' = 0$. The left hand side is a polynomial in K_h whose coefficients depend on M, M', T and T', and K_h is a root of this polynomial. Even though its degree can be quite high, this is often enough to mount devastating attacks. This weakness was one of the main criticism against the GCM authenticated encryption mode [MV04], whose authentication relies on the standard Wegman-Carter construction [Jou06].

The classical way to remedy this situation and achieve nonce-misuse resistance for Wegman-Carter MACs is to apply an extra PRF layer to the output of the construction. When this additional layer is a block cipher, one obtains what we call the *Encrypted Wegman-Carter* (EWC) construction. Let $F : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a keyed function, $E : \mathcal{K}' \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher, and $H : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ε-AXU hash function. Then the EWC construction based on F, E, and H has key space $\mathcal{K} \times \mathcal{K}' \times \mathcal{K}_h$, nonce space $\{0,1\}^n$, message space $\mathcal{M}$, and range $\{0,1\}^n$, and is defined by

$$\begin{aligned}\mathsf{EWC}[F,E,H]_{K,K',K_h}(N,M) &= E_{K'}\big(\mathsf{WC}[F,H]_{K,K_h}(N,M)\big)\\ &= E_{K'}\big(F_K(N)\oplus H_{K_h}(M)\big).\end{aligned}$$

One can straightforwardly verify that the security of this construction against nonce-respecting adversaries does not depend on E and that the upper bound of Theorem 1 still holds. For nonce-misusing adversaries, one has the following (the proof is omitted since it is exactly the same, *mutatis mutandis*, as the proof of Theorem 4 of Section 5).

Theorem 2. *Let F, E and H be as above. Then for any (q_m, q_v, t)-nonce-misusing adversary A against the MAC-security of $\mathsf{EWC}[F,E,H]$, there exists a (q_m+q_v, t')-adversary A' against the PRF-security of F and a (q_m+q_v, t'')-adversary A'' against the PRP-security of E, where $t', t'' = O(t+(q_m+q_v)t_H)$, such that*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\mathsf{EWC}[F,E,H]}(\mathsf{A}) \le \mathbf{Adv}^{\mathrm{PRF}}_{F}(\mathsf{A}') + \mathbf{Adv}^{\mathrm{PRP}}_{E}(\mathsf{A}'') + \frac{2(q_m+q_v)^2}{2^n} + \frac{(q_m+q_v)^2\varepsilon}{2}.$$

It is tempting to implement F from E. The simplest way to do so is simply to let $F = E$, thereby obtaining the construction (overloading notation $\mathsf{EWC}[\cdot]$)

$$\mathsf{EWC}[E,H]_{K,K',K_h}(N,M) = E_{K'}\big(E_K(N)\oplus H_{K_h}(M)\big).$$

However, the resulting MAC suffers from the same birthday-bound type problem against nonce-respecting adversaries as the unencrypted Wegman-Carter MAC $\mathsf{WC}[E,H]$ of Section 3.1. As already mentioned in introduction, it is possible to use a PRP-to-PRF conversion method to obtain security beyond the birthday bound, but using the best known constructions yields a MAC that makes at least three calls to the underlying block cipher. Our goal is to reduce the number of block cipher calls to two, which seems to be the minimum to achieve both security beyond the birthday bound and nonce-misuse resistance.

3.3 The New Construction EWCDM

The main contribution of this paper is to propose a much simpler solution that allows to get beyond the birthday bound, namely using the Davies-Meyer (DM) construction which turns a block cipher $E : \mathcal{K}\times\{0,1\}^n \to \{0,1\}^n$ into a keyed function as

$$\mathsf{DM}[E]_K(N) = E_K(N)\oplus N.$$

Using the DM construction based on E to instantiate F in $\mathsf{EWC}[F,E,H]$ results in a MAC construction based only on E and H, which we call *Encrypted Wegman-Carter with Davies-Meyer* (EWCDM) construction and denote $\mathsf{EWCDM}[E,H]$, illustrated on Fig. 1 and defined as follows:

$$\begin{aligned}\mathsf{EWCDM}[E,H]_{K,K',K_h}(N,M) &\stackrel{\text{def}}{=} \mathsf{EWC}[\mathsf{DM}[E],E,H]_{K,K',K_h}(N,M)\\ &= E_{K'}\big(E_K(N)\oplus N\oplus H_{K_h}(M)\big).\end{aligned}$$

As already explained in introduction, the DM construction is *not* PRF-secure beyond the birthday bound. Still, our main result, that we state and prove in the next section, is that the EWCDM construction is secure up to roughly $2^{2n/3}$ MAC queries and 2^n verification queries against nonce-respecting adversaries (while against nonce-misusing adversaries it still enjoys birthday-bound security).

The security proof entails an analysis of what we call the *Encrypted Davies-Meyer* (EDM) PRP-to-PRF conversion method, which turns two independent permutations P and P' of $\{0,1\}^n$ into a function of $\{0,1\}^n$ to $\{0,1\}^n$ defined as

$$\mathsf{EDM}[P, P'](N) = P'(P(N) \oplus N).$$

By "stripping off" from the security proof of EWCDM all details related to the hash function and verification queries, one can extract a proof that the EDM construction is a secure PRF up to $2^{2n/3}$ adversarial queries. We do so in Appendix A, and the reader might want to read this simpler proof before proceeding to Section 4. However, as already explained in introduction, it does not seem possible to prove the MAC-security of the EWCDM construction in a modular way from the PRF-security of the EDM construction.

Finally, note that adding the hash of the message to the output of the EDM construction (rather than "in the middle") would result in a construction secure up to $2^{2n/3}$ queries against nonce-respecting adversaries, but insecure against nonce-misusing ones since it is just an instantiation of the standard WC construction of Section 3.1 (with the EDM construction as PRF).

4 Nonce-Respecting Security of EWCDM

4.1 Statement of the Result and Overview of the Proof

In all the following, we simply denote $\Pi[E, H]$ the EWCDM construction based on block cipher E and AXU hash function H. Our main security result is as follows.

Theorem 3. *Let $\mathcal{M}$, $\mathcal{K}$ and $\mathcal{K}_h$ be non-empty sets. Let $E : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $H : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ε-AXU hash function. Then for any (q_m, q_v, t)-nonce-respecting adversary A against the MAC-security of $\Pi[E, H]$ with $q_m^{3/2} \leq 2^n/4$ and $q_v \leq 2^n/4$, there exists a $(q_m + q_v, t')$-adversary A' against the PRP-security of E, where $t' = O(t + (q_m + q_v)t_H)$, such that*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\Pi[E,H]}(\mathsf{A}) \leq 2\mathbf{Adv}^{\mathrm{PRP}}_{E}(\mathsf{A}') + \frac{5q_m^{3/2}}{2^n} + \frac{\varepsilon q_m}{2} + \frac{6q_v}{2^n} + \varepsilon q_v.$$

Hence, assuming $\varepsilon \simeq 2^{-n}$, the EWCDM construction is secure up to $q_m \simeq 2^{2n/3}$ MAC queries and $q_v \simeq 2^n$ verification queries.

In the remaining of the section, we prove Theorem 3. We fix a (q_m, q_v, t)-nonce-respecting adversary A against the MAC-security of $\Pi[E, H]$ and we let

$$\delta = \mathbf{Adv}^{\mathrm{MAC}}_{\Pi[E,H]}(\mathsf{A}).$$

As specified in Def. 1, adversary A has access to a MAC oracle $\Pi[E,H]_{K,K',K_h}$ and a verification oracle Ver_{K,K',K_h} for a randomly drawn key tuple (K,K',K_h).

The first step of the proof is standard and consists in replacing E_K and $E_{K'}$ by two random and independent permutations P and P', both in the MAC and in the verification oracle (in other words, we replace the block cipher E by the *perfect cipher* E^* whose key space is the set of all permutations of $\{0,1\}^n$). Let $\Pi[E^*,H]$ denote the resulting construction. It is easy to show that there exists an adversary against the PRP-security of E, making at most q_m+q_v oracle queries and runnig in time at most $O(t+(q_m+q_v)t_H)$, such that

$$\delta \leq 2\mathbf{Adv}_E^{\mathrm{PRP}}(\mathsf{A}') + \mathbf{Adv}_{\Pi[E^*,H]}^{\mathrm{MAC}}(\mathsf{A}). \tag{4}$$

(We replace successively E_K and $E_{K'}$ by a random permutation, each time constructing an hybrid PRP-adversary, and we consider the best of the two adversaries). Our goal is now to upper bound

$$\begin{aligned}\delta^* &\stackrel{\text{def}}{=} \mathbf{Adv}_{\Pi[E^*,H]}^{\mathrm{MAC}}(\mathsf{A}) \\ &= \Pr\left[(P,P') \leftarrow_{\$} \mathsf{Perm}(n)^2, K_h \leftarrow_{\$} \mathcal{K}_h : \mathsf{A}^{\Pi[P,P',H_{K_h}],\mathsf{Ver}[P,P',H_{K_h}]} \text{ forges}\right],\end{aligned}$$

where, overloading the notation, $\Pi[P,P',H_{K_h}]$ denotes the construction $\Pi[E^*,H]$ instantiated with permutations P, P', and hashing key K_h and $\mathsf{Ver}[P,P',H_{K_h}]$ denotes the corresponding verification oracle.

It will be more convenient to express δ^* as a *distinguishing* advantage. Namely, let Rand denote a perfectly random oracle with domain $\{0,1\}^n \times \mathcal{M}$ and range $\{0,1\}^n$, and Rej be an oracle with inputs in $\{0,1\}^n \times \mathcal{M} \times \{0,1\}^n$ which always returns 0 ("reject"). Since the adversary cannot forge (i.e., have the right oracle return 1) when interacting with $(\mathsf{Rand},\mathsf{Rej})$, we have

$$\delta^* = \Pr\left[\mathsf{A}^{\Pi[P,P',H_{K_h}],\mathsf{Ver}[P,P',H_{K_h}]} \text{ forges}\right] - \Pr\left[\mathsf{A}^{\mathsf{Rand},\mathsf{Rej}} \text{ forges}\right].$$

Consider now an adversary D which queries a pair of oracles $(\mathcal{O}_1,\mathcal{O}_2)$ and outputs a bit β, which we denote $\mathsf{D}^{\mathcal{O}_1,\mathcal{O}_2}=\beta$. (We will refer to such an adversary as a *distinguisher*.) Say that such an adversary is *non-trivial* if it never makes a query (N,M,T) to its right (verification) oracle if a previous query (N,M) to its left (MAC) oracle returned T. Then

$$\delta^* \leq \max_{\mathsf{D}} \Pr\left[\mathsf{D}^{\Pi[P,P',H_{K_h}],\mathsf{Ver}[P,P',H_{K_h}]}=1\right] - \Pr\left[\mathsf{D}^{\mathsf{Rand},\mathsf{Rej}}=1\right], \tag{5}$$

where the maximum is taken over non-trivial adversaries. (This follows easily by considering the particular D which runs A and outputs 1 iff A successfully forges.) Hence, we see that δ^* cannot be larger than the advantage of the best non-trivial distinguisher between the two pairs of oracles $(\Pi[P,P',H_{K_h}],\mathsf{Ver}[P,P',H_{K_h}])$ and $(\mathsf{Rand},\mathsf{Rej})$.[5] This formulation of the problem now allows us to use the H-coefficients technique [Pat08b, CS14], as we explain in more details below.

[5] While a verification query answered by 1 constitutes an obvious distinguishing criterion between the two worlds, a more advanced adversary might also use the small difference between the distributions of the answers of the left (MAC) oracle.

THE H-COEFFICIENTS TECHNIQUE. From now on, we fix a non-trivial distinguisher D interacting either with the *real world* $(\Pi[P, P', H_{K_h}], \mathsf{Ver}[P, P', H_{K_h}])$ for uniformly random permutations (P, P') and a random hashing key K_h, or with the *ideal world* $(\mathsf{Rand}, \mathsf{Rej})$, making at most q_m queries to its left (MAC) oracle and at most q_v queries to its right (verification) oracle, and outputting a single bit. We let

$$\mathbf{Adv}(\mathsf{D}) = \Pr\left[\mathsf{D}^{\Pi[P,P',H_{K_h}],\mathsf{Ver}[P,P',H_{K_h}]} = 1\right] - \Pr\left[\mathsf{D}^{\mathsf{Rand},\mathsf{Rej}} = 1\right].$$

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

$$\tau_m = \big((N_1, M_1, T_1), \ldots, (N_{q_m}, M_{q_m}, T_{q_m})\big)$$

be the list of MAC queries of D and corresponding answers. Let also

$$\tau_v = \big((N'_1, M'_1, T'_1, b_1), \ldots, (N'_{q_v}, M'_{q_v}, T'_{q_v}, b_{q_v})\big)$$

be the list of verification queries of D and corresponding answers (with $b_i \in \{0, 1\}$). The pair (τ_m, τ_v) constitutes the *queries transcript* of the attack. For convenience, we slightly modify the security experiment by revealing to the distinguisher (after it made all its queries but before it outputs its decision bit) the hashing key K_h if we are in the real world, or a uniformly random "dummy" key K_h if we are in the ideal world (this is obviously *wlog* since the distinguisher can ignore this additional piece of information). All in all, the *transcript* of the attack is the triplet $\tau = (\tau_m, \tau_v, K_h)$. We will often simply name a tuple $(N, M, T) \in \tau_m$ a *MAC query*, and a tuple $(N', M', T', b) \in \tau_v$ a *verification query*.

A transcript τ is said *attainable* (with respect to distinguisher D) if the probability to obtain this transcript in the ideal world is non-zero. In particular, note that for an attainable transcript $\tau = (\tau_m, \tau_v, K_h)$, any verification query $(N'_i, M'_i, T'_i, b_i) \in \tau_v$ is such that $b_i = 0$.[6] We denote Θ the set of attainable transcripts. We also denote X_{re}, resp. X_{id}, the probability distribution of the transcript τ induced by the real world, resp. the ideal world. The main lemma of the H-coefficients technique is the following one (see e.g. [CS14] or [CLL+14] for the proof).

Lemma 1. *Fix a distinguisher* D. *Let* $\Theta = \Theta_{\mathrm{good}} \sqcup \Theta_{\mathrm{bad}}$ *be a partition of the set of attainable transcripts. Assume that there exists* ε_1 *such that for any* $\tau \in \Theta_{\mathrm{good}}$, *one has*[7]

$$\frac{\Pr[X_{\mathrm{re}} = \tau]}{\Pr[X_{\mathrm{id}} = \tau]} \geq 1 - \varepsilon_1,$$

and that there exists ε_2 *such that* $\Pr[X_{\mathrm{id}} \in \Theta_{\mathrm{bad}}] \leq \varepsilon_2$. *Then* $\mathbf{Adv}(\mathsf{D}) \leq \varepsilon_1 + \varepsilon_2$.

---

6 Hence, some transcripts are attainable in the real world but not in the ideal world. While this is unusual (in most H-coefficients-based proofs, the set of transcripts attainable in the real world is a subset of those attainable in the ideal world), this is not a problem for Lemma 1 to hold.

7 Recall that for an attainable transcript, one has $\Pr[X_{\mathrm{id}} = \tau] > 0$.

The remaining of the proof of Theorem 3 is structured as follows: in Section 4.2, we define bad transcripts and upper bound their probability in the ideal world; in Section 4.3, we analyze good transcripts and prove that they are almost as likely in the real and the ideal world. Theorem 3 follows easily by combining Eqs. (4) and (5) above, Lemma 1, and Lemmas 2 and 3 proven below.

4.2 Definition and Probability of Bad Transcripts

We start by defining bad transcripts. We say that a MAC query $(N_i, M_i, T_i) \in \tau_m$ is *collisioning* if there exists another MAC query $(N_j, M_j, T_j) \in \tau_m$ with $j \neq i$ such $T_i = T_j$, otherwise we say it is *non-collisioning*.

Definition 3. *We say that an attainable transcript* $\tau = (\tau_m, \tau_v, K_h)$ *is* bad *if one of the following conditions is met:*

(i) the number of collisioning MAC queries in τ_m *is more than* $\sqrt{q_m}$*;*

(ii) there exists two distinct MAC queries (N_i, M_i, T_i) *and* (N_j, M_j, T_j) *in* τ_m *such that*

$$\begin{cases} T_i = T_j \\ N_i \oplus H_{K_h}(M_i) = N_j \oplus H_{K_h}(M_j); \end{cases}$$

(iii) there exists a MAC query $(N_i, M_i, T_i) \in \tau_m$ *and a verification query* $(N'_j, M'_j, T'_j, b_j) \in \tau_v$ *such that*

$$\begin{cases} N_i = N'_j \\ T_i = T'_j \\ H_{K_h}(M_i) = H_{K_h}(M'_j). \end{cases}$$

We denote Θ_{bad}*, resp.* Θ_{good} *the set of bad, respectively good transcripts.*

We quickly comment on these three conditions. Condition (i) captures the case where there are too many tag collisions and will be needed when lower bounding the probability of getting a good transcript in the real world. Condition (ii) can only happen in the ideal world and hence allows to trivially distinguish; in the real world, if $N_i \oplus H_{K_h}(M_i) = N_j \oplus H_{K_h}(M_j)$, then, since $N_i \neq N_j$ because the adversary is assumed nonce-respecting, one necessarily has

$$P(N_i) \oplus N_i \oplus H_{K_h}(M_i) \neq P(N_j) \oplus N_j \oplus H_{K_h}(M_j)$$

which implies $T_i \neq T_j$ by applying P' to both sides of the inequality. Similarly, condition (iii) can only happen in the ideal world since in the real world, if $N_i = N'_j$, $T_i = T'_j$, and $H_{K_h}(M_i) = H_{K_h}(M'_j)$, one should have $b_j = 1$ (while $b_j = 0$ in the ideal world).

We now upper bound the probability to get a bad transcript in the ideal world.

Lemma 2. *For any integers* q_m *and* q_v*, one has*

$$\Pr\left[X_{\text{id}} \in \Theta_{\text{bad}}\right] \leq \frac{q_m^{3/2}}{2^n} + \frac{\varepsilon q_m}{2} + \varepsilon q_v.$$

Proof. We upper bound the probabilities of the three conditions in turn. We denote Θ_i the set of attainable transcript that satisfy the i-th condition. Recall that, in the ideal world, K_h is drawn independently from the queries transcript.

Conditions (i) and (ii). We will deal with conditions (i) and (ii) together, using the fact that

$$\Pr[X_{\text{id}} \in \Theta_1 \vee X_{\text{id}} \in \Theta_2] \leq \Pr[X_{\text{id}} \in \Theta_1] + \Pr[X_{\text{id}} \in \Theta_2 \mid X_{\text{id}} \notin \Theta_1].$$

Since the adversary does not make useless queries, its MAC queries are distinct. In the ideal world, the values T_i for $i \in \{1, \ldots, q_m\}$ are then simply chosen uniformly and independently at random from $\{0,1\}^n$. We introduce the random variable

$$C = \left|\left\{(i,j) \in \{1, \ldots, q_m\}^2, i \neq j, T_i = T_j\right\}\right|.$$

The number of collisioning MAC queries is always lower than C. Note that

$$\mathbb{E}[C] = \sum_{1 \leq i \leq q_m} \sum_{\substack{1 \leq j \leq q_m \\ i \neq j}} \Pr[T_i = T_j] \leq \frac{q_m^2}{2^n}.$$

By Markov's inequality,

$$\Pr[X_{\text{id}} \in \Theta_1] \leq \Pr[C \geq \sqrt{q_m}] \leq \frac{q_m^{3/2}}{2^n}.$$

Assume now that $X_{\text{id}} \notin \Theta_1$, i.e., τ_m is such that the number of collisioning MAC queries is lower than $\sqrt{q_m}$. Recall that K_h is chosen independently from τ_m in the ideal world. Fix any (i,j) such that $i \neq j$ and $T_i = T_j$. Since the number of collisioning MAC queries is lower than $\sqrt{q_m}$, there are at most $q_m/2$ such pairs of queries. Then, since H is ε-AXU, one has

$$\Pr[K_h \leftarrow_{\$} \mathcal{K}_h : N_i \oplus H_{K_h}(M_i) = N_j \oplus H_{K_h}(M_j)] \leq \varepsilon$$

and, by summing over the at most $q_m/2$ such pairs of queries, one has

$$\Pr[X_{\text{id}} \in \Theta_2 \mid X_{\text{id}} \notin \Theta_1] \leq \frac{\varepsilon q_m}{2}.$$

Hence,

$$\Pr[X_{\text{id}} \in \Theta_1 \cup \Theta_2] \leq \frac{q_m^{3/2}}{2^n} + \frac{\varepsilon q_m}{2}.$$

Condition (iii). We consider any verification query $(N'_j, M'_j, T'_j, b_j) \in \tau_v$ and upper bound the probability that condition (iii) is satisfied for this particular query. Since the adversary is nonce-respecting, there is at most one MAC query (N_i, M_i, T_i) such that $N_i = N'_j$. We distinguish two cases:

– If the verification query comes after the MAC query, then since the distinguisher is non-trivial, either $T_i \neq T'_j$, or $M_i \neq M'_j$. In the former case, condition (iii) cannot be satisfied, while in the latter case, the probability over the random draw of K_h that $H_{K_h}(M_i) \oplus H_{K_h}(M'_j) = 0$ is at most ε.

– If the MAC query comes after the verification query, then T_i is random and independent from T'_j and the probability that $T_i = T'_j$ is 2^{-n}.

Since for an ε-AXU hash function with range $\{0,1\}^n$ one has $\varepsilon \geq 2^{-n}$, we see that in all cases condition (iii) is met with probability at most ε. Thus, by summing over every verification query, one has

$$\Pr\left[X_{\mathrm{id}} \in \Theta_3\right] \leq \varepsilon q_v.$$

The lemma follows by an union bound over all conditions. □

4.3 Analysis of Good Transcripts

We now analyze good transcripts and prove the following lemma.

Lemma 3. *Assume that $q_m^{3/2} \leq 2^n/4$ and $q_v \leq 2^n/4$. Then, for any good transcript τ, one has*

$$\frac{\Pr\left[X_{\mathrm{re}} = \tau\right]}{\Pr\left[X_{\mathrm{id}} = \tau\right]} \geq 1 - \frac{4q_m^{3/2}}{2^n} - \frac{6q_v}{2^n}.$$

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

$$\Pr[X_{\mathrm{id}} = \tau] = \frac{1}{|\mathcal{K}_h| \cdot (2^n)^{q_m}}. \tag{6}$$

We must now lower bound the probability of getting τ in the real world. We say that a pair of permutations (P, P') is compatible with τ_m if

$$\forall i \in \{1, \ldots, q_m\},\ \Pi[P, P', H_{K_h}](N_i, M_i) = T_i,$$

and we say that it is compatible with τ_v if

$$\forall i \in \{1, \ldots, q_v\},\ \Pi[P, P', H_{K_h}](N'_i, M'_i) \neq T'_i.$$

We simply say that (P, P') is compatible with τ if it is compatible with τ_m and τ_v. We denote $\mathsf{Comp}(\tau_m)$, $\mathsf{Comp}(\tau_v)$, and $\mathsf{Comp}(\tau)$ the set of pairs of permutations that are compatible with respectively τ_m, τ_v, and τ. Then one can easily check (see for example [CS14] for a detailed explanation) that

$$\Pr[X_{\mathrm{re}} = \tau] = \frac{1}{|\mathcal{K}_h|} \cdot \Pr\left[(P, P') \leftarrow_{\$} \mathsf{Perm}(n)^2 : (P, P') \in \mathsf{Comp}(\tau)\right]. \tag{7}$$

MAC Queries Transcript. We will first consider the probability that a random pair (P, P') is compatible with the MAC queries transcript τ_m. To ease the notation, we reorder the transcript as follows. Let r be the number of distinct tags T appearing in MAC queries. Then we rewrite the transcript so that all queries with the same tag are consecutive, so that the MAC queries transcript (that we still denote τ_m) is now

$$\begin{aligned}\tau_m = \big(&(N_{1,1}, M_{1,1}, T_1), \ldots, (N_{1,q_1}, M_{1,q_1}, T_1),\\ &(N_{2,1}, M_{2,1}, T_2), \ldots, (N_{2,q_2}, M_{2,q_2}, T_2),\\ &\ldots,\\ &(N_{r,1}, M_{r,1}, T_r), \ldots, (N_{r,q_r}, M_{r,q_r}, T_r)\big),\end{aligned}$$

where $T_1, \ldots, T_r$ are distinct and $\sum_{i=1}^{r} q_i = q_m$.

Our goal is now to lower bound the probability that a random pair of permutations (P, P') satisfies

$$\forall i \in \{1, \ldots, r\}, \forall j \in \{1, \ldots, q_i\},\ P'\big(P(N_{i,j}) \oplus N_{i,j} \oplus H_{K_h}(M_{i,j})\big) = T_i.$$

For this, we will consider the possible "internal" values $Z_i = (P')^{-1}(T_i)$. We say that a tuple $\mathbf{Z} = (Z_1, \ldots, Z_r)$ of distinct values in $\{0,1\}^n$ is *good* if

- (a) all q_m values $Z_i \oplus N_{i,j} \oplus H_{K_h}(M_{i,j})$ for $i \in \{1, \ldots, r\}$, $j \in \{1, \ldots, q_i\}$ are distinct;
- (b) for every verification query $(N', M', T', b) \in \tau_v$ such that $N' = N_{i,j}$ and $T' = T_k$ for some $i \in \{1, \ldots, r\}$, $j \in \{1, \ldots, q_i\}$, and $k \in \{1, \ldots, r\}$ with $k \neq i$, one has
$$Z_i \oplus H_{K_h}(M_{i,j}) \oplus H_{K_h}(M') \neq Z_k.$$

Property (a) is needed since the values $Z_i \oplus N_{i,j} \oplus H_{K_h}(M_{i,j})$ are the images by P of the (distinct) nonces $N_{i,j}$. Property (b) will be needed later when lower bounding the probability that (P, P') is compatible with the verification transcript τ_v.

Given a good tuple $\mathbf{Z}$, the probability, for a randomly drawn pair (P, P'), that

$$\begin{cases} \forall i \in \{1, \ldots, r\}, \forall j \in \{1, \ldots, q_i\},\ P(N_{i,j}) = Z_i \oplus N_{i,j} \oplus H_{K_h}(M_{i,j}), \\ \forall i \in \{1, \ldots, r\},\ P'(Z_i) = T_i \end{cases} \tag{8}$$

is exactly

$$\frac{1}{(2^n)_{q_m}(2^n)_r}. \tag{9}$$

(This is simply the probability that P satisfies $q_1 + \ldots + q_r = q_m$ equations and P' satisfies r equations.)

It remains to lower bound the number $N_{\mathbf{Z}}$ of good tuples $\mathbf{Z}$, which can be done as follows. First, note that by definition of a good transcript, for any $i \in \{1, \ldots, r\}$, the values $Z_i \oplus N_{i,j} \oplus H_{K_h}(M_{i,j})$ for $1 \leq j \leq q_i$ are distinct since otherwise condition (ii) defining a bad transcript would be fulfilled (without that, good tuples $\mathbf{Z}$ would not exist). In the following, for $i, k \in \{1, \ldots, r\}$ with $k < i$, we denote $q'_{i,k}$ the number of verification queries $(N', M', T', b) \in \tau_v$ such that either $N' = N_{i,j}$ for some $j \in \{1, \ldots, q_i\}$ and $T' = T_k$, or $N' = N_{k,j}$ for some

$j \in \{1, \ldots, q_k\}$ and $T' = T_i$. Note that since a verification query can count for at most one pair (i, k), one has

$$\sum_{i=2}^{r} \sum_{k=1}^{i-1} q'_{i,k} \leq q_v. \tag{10}$$

Then,

- there are at least 2^n possibilities for Z_1;
- once Z_1 is fixed, there are at least $2^n - 1 - q_2 q_1 - q'_{2,1}$ possibilities for Z_2 since Z_2 must be different from the following values:
  - Z_1,
  - $Z_1 \oplus N_{1,j} \oplus H_{K_h}(M_{1,j}) \oplus N_{2,j'} \oplus H_{K_h}(M_{2,j'})$ for all $j \in \{1, \ldots, q_1\}$ and all $j' \in \{1, \ldots, q_2\}$ (in order for property (a) to be fulfilled),
  - $Z_1 \oplus H_{K_h}(M_{1,j}) \oplus H_{K_h}(M')$ for every verification query $(N', M', T', b) \in \tau_v$ such that $N' = N_{1,j}$ for some $j \in \{1, \ldots, q_1\}$ and $T' = T_2$, and $Z_1 \oplus H_{K_h}(M_{2,j}) \oplus H_{K_h}(M')$ for every verification query $(N', M', T', b) \in \tau_v$ such that $N' = N_{2,j}$ for some $j \in \{1, \ldots, q_2\}$ and $T' = T_1$, which amounts to at most $q'_{2,1}$ values (in order for property (b) to be fulfilled);
- once $Z_1, \ldots, Z_i$ are fixed, there are at least $2^n - i - q_{i+1} \sum_{k=1}^{i} q_k - \sum_{k=1}^{i} q'_{i+1,k}$ possibilities for Z_{i+1} since Z_{i+1} must be different from the following values:
  - $Z_1, \ldots, Z_i$,
  - $Z_k \oplus N_{k,j} \oplus H_{K_h}(M_{k,j}) \oplus N_{i+1,j'} \oplus H_{K_h}(M_{i+1,j'})$ for all $k \in \{1, \ldots, i\}$, all $j \in \{1, \ldots, q_k\}$, and all $j' \in \{1, \ldots, q_{i+1}\}$,
  - $Z_k \oplus H_{K_h}(M_{k,j}) \oplus H_{K_h}(M')$ for every verification query $(N', M', T', b) \in \tau_v$ such that $N' = N_{k,j}$ for some $k \in \{1, \ldots, i\}$, $j \in \{1, \ldots, q_k\}$ and $T' = T_{i+1}$, and $Z_k \oplus H_{K_h}(M_{i+1,j}) \oplus H_{K_h}(M')$ for every verification query $(N', M', T', b) \in \tau_v$ such that $N' = N_{i+1,j}$ for some $j \in \{1, \ldots, q_{i+1}\}$ and $T' = T_k$ for some $k \in \{1, \ldots, i\}$, which amounts to at most $\sum_{k=1}^{i} q'_{i+1,k}$ values.

Hence, the number of good tuples $\mathbf{Z} = (Z_1, \ldots, Z_r)$ is at least

$$N_{\mathbf{Z}} \geq \prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{k=1}^{i} q_k - \sum_{k=1}^{i} q'_{i+1,k} \right). \tag{11}$$

Verification Queries Transcript. From now on, we fix a good tuple $\mathbf{Z}$. We will now lower bound the probability that a random pair (P, P') is compatible with the verification transcript τ_v, conditioned on (P, P') satisfying the set of equations (8). (Recall that P is then fixed on q_m values and P' is fixed on r values.) For this, it will be easier to upper bound the probability that (P, P') is *not* compatible with τ_v, i.e., that there exists $(N', M', T', b) \in \tau_v$ such that

$$P'\big(P(N') \oplus N' \oplus H_{K_h}(M')\big) = T'. \tag{12}$$

Fix any verification query $(N', M', T', b) \in \tau_v$. We say that it is *nonce-fresh*, resp. *tag-fresh*, if N', resp. T' does not appear in the MAC queries transcript τ_m.[8] We consider four possible cases.

– *Case 1: the verification query is both nonce-fresh and tag-fresh.* Then $P(N')$ is random and two sub-cases can occur: if $P(N') \oplus N' \oplus H_{K_h}(M') \in \mathbf{Z}$, Eq. (12) cannot be satisfied since the query is tag-fresh; on the other hand, if $P(N') \oplus N' \oplus H_{K_h}(M') \notin \mathbf{Z}$, Eq. (12) is satisfied with probability $1/(2^n - r)$ over the choice of P'. Hence, over the choice of (P, P'), Eq. (12) is satisfied with probability at most
$$\frac{1}{2^n - r} \leq \frac{1}{2^n - q_m}.$$
– *Case 2: the verification query is nonce-fresh, but not tag-fresh.* Then there exists $(N, M, T) \in \tau_m$ such that $T = T'$. Let $Z = (P')^{-1}(T)$ (this value is well defined since we assume Eqs. (8) are satisfied). Then Eq. (12) is satisfied iff
$$P(N') = Z \oplus N' \oplus H_{K_h}(M'),$$
hence with probability exactly $1/(2^n - q_m)$ since the query is nonce-fresh and N' does not appear in Eqs. (8).
– *Case 3: the verification query is tag-fresh, but not nonce-fresh.* Then there exists a unique $(N, M, T) \in \tau_m$ such that $N' = N$, so that $P(N')$ is fixed by Eqs. (8). If $P(N') \oplus N' \oplus H_{K_h}(M') \in \mathbf{Z}$, then Eq. (12) cannot be satisfied since the query is tag-fresh. If $P(N') \oplus N' \oplus H_{K_h}(M') \notin \mathbf{Z}$, then Eq. (12) is satisfied with probability
$$\frac{1}{2^n - r} \leq \frac{1}{2^n - q_m}.$$
– *Case 4: the verification query is neither nonce-fresh nor tag-fresh.* Then there exists a unique $(N_{i,j}, M_{i,j}, T_i) \in \tau_m$ such that $N' = N_{i,j}$ and $(N_k, M_k, T_k) \in \tau_m$ (with possibly $k = i$) such that $T' = T_k$. If $k = i$, then Eq. (12) cannot be satisfied since otherwise one would have
$$P(N') \oplus N' \oplus H_{K_h}(M') = (P')^{-1}(T_i) = P(N_{i,j}) \oplus N_{i,j} \oplus H_{K_h}(M_{i,j}),$$
which implies $H_{K_h}(M') = H_{K_h}(M_{i,j})$ and condition (*iii*) defining a bad transcript would be fulfilled. On the other hand, if $k \neq i$, then Eq. (12) being satisfied would imply
$$\begin{aligned} & P(N') \oplus N' \oplus H_{K_h}(M') = (P')^{-1}(T_k) = Z_k \\ \Rightarrow\ & P(N_{i,j}) \oplus N_{i,j} \oplus H_{K_h}(M') = Z_k \\ \Rightarrow\ & Z_i \oplus H_{K_h}(M_{i,j}) \oplus H_{K_h}(M') = Z_k, \end{aligned}$$

[8] We stress that this freshness definition is with respect to the entire MAC queries transcript τ_m, independently of when the verification query was actually made by the distinguisher.

and this would contradict property (b) of a good tuple $\mathbf{Z}$. Hence, by definition of a good transcript and a good tuple $\mathbf{Z}$, we see that Eq. (12) cannot be satisfied in that case.

Summarizing, we see that for any verification query, Eq. (12) is satisfied with probability at most $1/(2^n - q_m)$. By a union bound over the q_v verification queries, we obtain that

$$\Pr\left[(P, P') \in \mathsf{Comp}(\tau_v) \mid (P, P') \text{ satisfies Eqs. (8)}\right] \geq 1 - \frac{q_v}{2^n - q_m}. \tag{13}$$

Summing Up. We can now lower bound the probability that a random pair (P, P') is compatible with τ, that we denote

$$\mathsf{p}(\tau) \stackrel{\text{def}}{=} \Pr\left[(P, P') \leftarrow_\$ \mathsf{Perm}(n)^2 : (P, P') \in \mathsf{Comp}(\tau)\right].$$

Namely, summing over all good tuples $\mathbf{Z}$, and using (9), (11), and (13), we have

$$\begin{aligned}\mathsf{p}(\tau) &\geq N_{\mathbf{Z}} \times \Pr\left[(P, P') \text{ satisfies Eqs. (8)}\right] \\ &\quad \times \Pr\left[(P, P') \in \mathsf{Comp}(\tau_v) \mid (P, P') \text{ satisfies Eqs. (8)}\right] \\ &\geq \frac{\prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{k=1}^{i} q_k - \sum_{k=1}^{i} q'_{i+1,k}\right)}{(2^n)_{q_m} (2^n)_r} \left(1 - \frac{q_v}{2^n - q_m}\right).\end{aligned}$$

This, in turn, allows us to lower bound the ratio of the probabilities to obtain τ in the real and the ideal world, namely combining (6) and (7) with the equation above, we have

$$\begin{aligned}\frac{\Pr\left[X_{\text{re}} = \tau\right]}{\Pr\left[X_{\text{id}} = \tau\right]} &\geq \underbrace{\frac{(2^n)^{q_m} \prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{k=1}^{i} q_k - \sum_{k=1}^{i} q'_{i+1,k}\right)}{(2^n)_{q_m} (2^n)_r}}_{A} \\ &\quad \times \left(1 - \frac{q_v}{2^n - q_m}\right).\end{aligned} \tag{14}$$

We focus on term A, that we can rewrite

$$A = \prod_{i=0}^{q_m - 1} \left(1 + \frac{i}{2^n - i}\right) \prod_{i=0}^{r-1} \left(1 - \underbrace{\frac{q_{i+1} \sum_{k=1}^{i} q_k}{2^n - i}}_{a_i} - \underbrace{\frac{\sum_{k=1}^{i} q'_{i+1,k}}{2^n - i}}_{b_i}\right). \tag{15}$$

The following “Bonferroni-type” inequality will be useful to further lower bound A.

Lemma 4. *Let $r \geq 1$ be an integer and $(a_i)_{0 \leq i \leq r-1}$ and $(b_i)_{0 \leq i \leq r-1}$ be positive reals such that $a_i \leq 1/2$ and $b_i \leq 1/2$ for all $i \in \{0, \dots, r-1\}$. Then*

$$\prod_{i=0}^{r-1} (1 - a_i - b_i) \geq \prod_{i=0}^{r-1} (1 - a_i) \prod_{i=0}^{r-1} (1 - 2b_i).$$

Proof. The proof is by induction. We first prove it for $r = 1$. One has

$$(1-a_0)(1-2b_0) = 1 - a_0 - 2b_0 + 2a_0b_0 = 1 - a_0 - b_0 - \underbrace{b_0(1-2a_0)}_{\geq 0} \leq 1 - a_0 - b_0.$$

Assume that the result holds for $r \geq 1$. Then

$$\begin{aligned}
\prod_{i=0}^{r}(1-a_i)\prod_{i=0}^{r}(1-2b_i) &= \prod_{i=0}^{r-1}(1-a_i)\prod_{i=0}^{r-1}(1-2b_i) \times \underbrace{(1-a_r)(1-2b_r)}_{\geq 0} \\
&\leq \prod_{i=0}^{r-1}(1-a_i-b_i) \times (1 - a_r - b_r - b_r(1-2a_r)) \\
&= \prod_{i=0}^{r}(1-a_i-b_i) - \underbrace{b_r(1-2a_r)\prod_{i=0}^{r-1}(1-a_i-b_i)}_{\geq 0} \\
&\leq \prod_{i=0}^{r}(1-a_i-b_i).
\end{aligned}$$

The result holds for $r+1$ and the lemma follows. □

We can apply this lemma to the r.h.s. of (15). Indeed, for any $i \in \{0, \dots, r-1\}$, one has $q_{i+1} \leq \sqrt{q_m}$ (as otherwise condition (i) of a bad transcript would be met), and $q_m^{3/2} \leq 2^n/4$ by assumption, so that

$$a_i \stackrel{\text{def}}{=} \frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n - i} \leq \frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n - q_m} \leq \frac{2q_m^{3/2}}{2^n} \leq \frac{1}{2},$$

Moreover, by (10) and the assumption that $q_v \leq 2^n/4$, one has

$$b_i \stackrel{\text{def}}{=} \frac{\sum_{k=1}^{i} q'_{i+1,k}}{2^n - i} \leq \frac{\sum_{k=1}^{i} q'_{i+1,k}}{2^n - q_m} \leq \frac{2q_v}{2^n} \leq \frac{1}{2}.$$

Hence,

$$\begin{aligned}
A &\geq \prod_{i=0}^{q_m-1}\left(1 + \frac{i}{2^n - i}\right)\prod_{i=0}^{r-1}\left(1 - \frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n - i}\right)\prod_{i=0}^{r-1}\left(1 - \frac{2\sum_{k=1}^{i} q'_{i+1,k}}{2^n - i}\right) \\
&\geq \prod_{i=0}^{q_m-1}\left(1 + \frac{i}{2^n - i}\right)\prod_{i=0}^{r-1}\left(1 - \frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n - i}\right)\left(1 - \frac{2\sum_{i=0}^{r-1}\sum_{k=1}^{i} q'_{i+1,k}}{2^n - q_m}\right) \\
&\geq \underbrace{\prod_{i=0}^{q_m-1}\left(1 + \frac{i}{2^n - i}\right)\prod_{i=0}^{r-1}\left(1 - \frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n - i}\right)}_{A'}\left(1 - \frac{2q_v}{2^n - q_m}\right), \qquad (16)
\end{aligned}$$

where for the last inequality we used (10).

In order to further lower bound A', we need to distinguish collisioning MAC queries from non-collisioning ones. Up to reordering the MAC queries transcript, we assume that non-collisioning queries come first, and we let $s \in \{0, \dots, r\}$ be the integer such that $q_i = 1$ for $i \in \{1, \dots, s\}$, and $q_i > 1$ for $i \in \{s+1, \dots, r\}$. Note that since the transcript is good, one has

$$\sum_{i=s+1}^{r} q_i \leq \sqrt{q_m} \tag{17}$$

as otherwise condition (i) of a bad transcript would be fulfilled. Then

$$\begin{aligned}
A' &\geq \prod_{i=0}^{q_m-1}\left(1+\frac{i}{2^n-i}\right)\prod_{i=0}^{s-1}\left(1-\frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n-i}\right)\prod_{i=s}^{r-1}\left(1-\frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n-i}\right)\\
&= \prod_{i=0}^{q_m-1}\left(1+\frac{i}{2^n-i}\right)\prod_{i=0}^{s-1}\left(1-\frac{i}{2^n-i}\right)\prod_{i=s}^{r-1}\left(1-\frac{q_{i+1}\sum_{k=1}^{i} q_k}{2^n-i}\right)\\
&\geq \prod_{i=0}^{q_m-1}\left(1-\frac{i^2}{(2^n-i)^2}\right)\prod_{i=s}^{r-1}\left(1-\frac{q_{i+1}q_m}{2^n-i}\right)\\
&\geq \prod_{i=0}^{q_m-1}\left(1-\frac{i^2}{(2^n-q_m)^2}\right)\prod_{i=s}^{r-1}\left(1-\frac{q_{i+1}q_m}{2^n-q_m}\right)\\
&\geq \left(1-\frac{q_m^3}{3(2^n-q_m)^2}\right)\left(1-\frac{q_m\sum_{i=s+1}^{r} q_i}{2^n-q_m}\right)\\
&\geq \left(1-\frac{4q_m^3}{3\cdot 2^{2n}}\right)\left(1-\frac{2q_m^{3/2}}{2^n}\right),
\end{aligned} \tag{18}$$

where for the last inequality we used (17) and $q_m \leq 2^n/2$.

Combining (14), (16), and (18), we finally obtain (using $q_m \leq 2^n/2$ once again)

$$\frac{\Pr\left[X_{\mathrm{re}}=\tau\right]}{\Pr\left[X_{\mathrm{id}}=\tau\right]} \geq 1-\frac{4q_m^3}{3\cdot 2^{2n}}-\frac{2q_m^{3/2}}{2^n}-\frac{6q_v}{2^n}.$$

Lemma 3 follows using $q_m^3/2^{2n} \leq q_m^{3/2}/2^n$ by our assumption that $q_m^{3/2} \leq 2^n/4$.

5 Nonce-Misuse Security of EWCDM

In this section, we consider the security of the EWCDM construction when the adversary is allowed to repeat nonces. In this setting, PRF-security implies MAC-security, hence we can simply consider the EWCDM construction as a function with domain $\mathcal{N} \times \mathcal{M}$ and study its pseudorandomness. Our result on the PRF-security of the EWCDM construction is as follows.

Lemma 5. *Let $\mathcal{M}$, $\mathcal{K}$ and $\mathcal{K}_h$ be non-empty sets. Let $E : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $H : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ε-AXU hash function. Then for any (q,t)-(nonce-misusing) adversary A against the PRF-security of $\Pi[E,H]$, there exists a (q,t')-adversary A' against the PRP-security of E, where $t' = O(t + qt_H)$, such that*

$$\mathbf{Adv}^{\mathrm{PRF}}_{\Pi[E,H]}(\mathsf{A}) \leq 2\mathbf{Adv}^{\mathrm{PRP}}_{E}(\mathsf{A}') + \frac{q^2}{2^n} + \frac{q^2\varepsilon}{2}.$$

The corresponding MAC-security can easily be deduced from Lemma 5 using the following generic result of Bellare *et al.* [BGM04, Proposition 7.3].

Lemma 6. *Let F be a keyed function with output length n. Then for any (q_m, q_v, t)-adversary A against the MAC-security of F, there exists a $(q_m + q_v, t')$-adversary A' against the PRF-security of F, where $t' = O(t)$, such that*

$$\mathbf{Adv}^{\mathrm{MAC}}_{F}(\mathsf{A}) \leq \mathbf{Adv}^{\mathrm{PRF}}_{F}(\mathsf{A}') + \frac{q_v}{2^n}.$$

Combining Lemmas 5 and 6, we obtain the following theorem (absorbing the $q_v/2^n$ term into $(q_m + q_v)^2/2^n$).

Theorem 4. *Let $\mathcal{M}$, $\mathcal{K}$ and $\mathcal{K}_h$ be non-empty sets. Let $E : \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$ be a block cipher and $H : \mathcal{K}_h \times \mathcal{M} \to \{0,1\}^n$ be an ε-AXU hash function. Then for any (q_m, q_v, t)-nonce-misusing adversary A against the MAC-security of $\Pi[E,H]$, there exists a $(q_m + q_v, t')$-adversary A' against the PRP-security of E, where $t' = O(t + (q_m + q_v)t_H)$, such that*

$$\mathbf{Adv}^{\mathrm{MAC}}_{\Pi[E,H]}(\mathsf{A}) \leq 2\mathbf{Adv}^{\mathrm{PRP}}_{E}(\mathsf{A}') + \frac{2(q_m + q_v)^2}{2^n} + \frac{(q_m + q_v)^2\varepsilon}{2}.$$

The proof of Lemma 5 is standard (indeed, the construction, seen as a keyed function with domain $\mathcal{N} \times \mathcal{M}$, follows the classical "hash-then-PRF" paradigm). We include it below for completeness.

Proof of Lemma 5. Fix a (q,t)-adversary A against the PRF-security of $\Pi[E,H]$. The first step of the proof consists in replacing E_K and $E_{K'}$ by two uniformly random and independent permutations P and P'. It is easy to show that there is an adversary A' making at most q queries and running in time at most $t' = O(t + qt_H)$ such that

$$\mathbf{Adv}^{\mathrm{PRF}}_{\Pi[E,H]}(\mathsf{A}) \leq 2\mathbf{Adv}^{\mathrm{PRP}}_{E}(\mathsf{A}') + \mathbf{Adv}^{\mathrm{PRF}}_{\Pi[E^*,H]}(\mathsf{A}), \tag{19}$$

where E^* denotes the perfect cipher on $\{0,1\}^n$. Then, we use the PRP/PRF switching lemma [BR06] to replace the random permutations P and P' by two independent and uniformly random functions R and R', obtaining

$$\mathbf{Adv}^{\mathrm{PRF}}_{\Pi[E^*,H]}(\mathsf{A}') \leq \frac{q^2}{2^n} + \mathbf{Adv}^{\mathrm{PRF}}_{\Pi[F^*,H]}(\mathsf{A}), \tag{20}$$

where F^* denotes the perfect keyed function from $\{0,1\}^n$ to $\{0,1\}^n$ (i.e., the keyed function with key space $\mathsf{Func}(n)$).

It remains to upper bound the PRF-advantage of A against $\Pi[F^*, H]$. For this, we use the H-coefficients technique. The adversary must distinguish between two worlds:

- the real world in which it interacts with $\Pi[R, R', H]$ where R and R' are two uniformly and independently drawn functions from $\{0,1\}^n$ to $\{0,1\}^n$;
- the ideal world in which it receives independent and uniformly random answers.

Let $\tau_m = ((N_1, M_1, T_1), \ldots, (N_q, M_q, T_q))$ be the list of all queries of A and the corresponding answers. In order to have a simple description of bad transcripts, we reveal to the adversary at the end of the experiment the key K_h and the function R if we are in the real world, while in the ideal world we simply draw a dummy key $K_h \leftarrow_\$ \mathcal{K}_h$ and a function R independently from the answers of the oracle. All in all, the transcript of the interaction of A with its oracle is a tuple $\tau = (\tau_m, K_h, R)$ and, in this case, a transcript is said attainable (with respect to an adversary A) if the probability to obtain it in the ideal world is non-zero. We denote Θ the set of attainable transcripts. We also denote X_{re}, resp. X_{id}, the probability distribution of the transcript τ induced by the real world, resp. the ideal world.

We start by defining the set of bad transcripts.

Definition 4. *We say that an attainable transcript $\tau = (\tau_m, K_h, R)$ is bad if there exists distinct queries $(N, M, T), (N', M', T') \in \tau_m$ such that*

$$R(N) \oplus N \oplus H_{K_h}(M) = R(N') \oplus N' \oplus H_{K_h}(M').$$

Otherwise we say that τ is good. We denote Θ_{bad}, resp. Θ_{good}, the set of bad, resp. good transcripts.

We first upper bound the probability to get a bad transcript in the ideal world.

Lemma 7.

$$\Pr\left[X_{\mathrm{id}} \in \Theta_{\mathrm{bad}}\right] \leq \frac{q^2 \varepsilon}{2}.$$

Proof. Let τ_m be any attainable query transcript. Recall that, in the ideal world, the key K_h and the function R are drawn uniformly at random and independently from the query transcript τ_m. Fix any pair of distinct queries $(N, M, T), (N', M', T')$. Two cases can occur:

- $M \neq M'$: then the probability, over the random draw of K_h and R, that $R(N) \oplus N \oplus H_{K_h}(M) = R(N') \oplus N' \oplus H_{K_h}(M')$ is lower than ε by the ε-AXU property of H;
- $M = M'$: then, since we assume that the adversary never makes redundant queries, $N \neq N'$ and the probability that $R(N) \oplus N = R(N') \oplus N'$ is $1/2^n \leq \varepsilon$.

By summing over every possible pair of queries, one gets the result. □

We then analyze good transcripts.

Lemma 8. *For any good transcript τ, one has*

$$\frac{\Pr[X_{\text{re}} = \tau]}{\Pr[X_{\text{id}} = \tau]} = 1.$$

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

$$\Pr[X_{\text{id}} = \tau] = \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{|\mathsf{Func}(n)|} \cdot \frac{1}{(2^n)^q}$$

since, in the ideal world, the oracle is perfectly random and the key K_h and the function R are chosen uniformly at random and independently from the query transcript.

We say that a function $R' \in \mathsf{Func}(n)$ is compatible with the transcript τ if $R'(R(N_i) \oplus N_i \oplus H_{K_h}(M_i)) = T_i$ for all $i \in \{1, \ldots, q\}$. Let $\mathsf{Comp}(\tau)$ be the set of all compatible functions R'. Then it is easy to see that

$$\Pr[X_{\text{re}} = \tau] = \frac{1}{|\mathcal{K}_h|} \cdot \frac{1}{|\mathsf{Func}(n)|} \cdot \Pr[R' \leftarrow_\$ \mathsf{Func}(n) : R' \in \mathsf{Comp}(\tau)].$$

Since τ is a good transcript, the values $R(N_i) \oplus N_i \oplus H_{K_h}(M_i)$ are distinct. Hence

$$\Pr[R' \leftarrow_\$ \mathsf{Func}(n) : R' \in \mathsf{Comp}(\tau)] = \frac{1}{(2^n)^q}$$

and therefore $\Pr[X_{\text{re}} = \tau] = \Pr[X_{\text{id}} = \tau]$. □

Combining Lemmas 1, 7, and 8, one obtains

$$\mathbf{Adv}^{\text{PRF}}_{\Pi[F^*,H]}(\mathsf{A}) \leq \frac{q^2 \varepsilon}{2}. \tag{21}$$

Lemma 5 finally follows from Eqs. (19), (20), and (21).

Acknowledgments

Many thanks to Thomas Peyrin. This paper stemmed from discussions with him, and he took part to the early stages of this work.

References

[ABBT15] Mohamed Ahmed Abdelraheem, Peter Beelen, Andrey Bogdanov, and Elmar Tischhauser. Twisted Polynomials and Forgery Attacks on GCM. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015 (Proceedings, Part I)*, volume 9056 of *LNCS*, pages 762–786. Springer, 2015.

[BC09] John Black and Martin Cochran. MAC Reforgeability. In Orr Dunkelman, editor, *Fast Software Encryption - FSE 2009*, volume 5665 of *LNCS*, pages 345–362. Springer, 2009.

[BCK96] Mihir Bellare, Ran Canetti, and Hugo Krawczyk. Keying Hash Functions for Message Authentication. In Neal Koblitz, editor, *Advances in Cryptology - CRYPTO '96*, volume 1109 of *LNCS*, pages 1–15. Springer, 1996.

[Ber00] Daniel J. Bernstein. Floating-Point Arithmetic and Message Authentication. Unpublished manuscript, 2000. Available at http://cr.yp.to/papers.html#hash127.

[Ber05a] Daniel J. Bernstein. Stronger Security Bounds for Permutations. Unpublished manuscript, 2005. Available at http://cr.yp.to/papers.html#permutations.

[Ber05b] Daniel J. Bernstein. Stronger Security Bounds for Wegman-Carter-Shoup Authenticators. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT 2005*, volume 3494 of *LNCS*, pages 164–180. Springer, 2005.

[Ber05c] Daniel J. Bernstein. The Poly1305-AES Message-Authentication Code. In Henri Gilbert and Helena Handschuh, editors, *Fast Software Encryption - FSE 2005*, volume 3557 of *LNCS*, pages 32–49. Springer, 2005.

[Ber07] Daniel J. Bernstein. Polynomial Evaluation and Message Authentication. Unpublished manuscript, 2007. Available at http://cr.yp.to/papers.html#pema.

[BGK99] Mihir Bellare, Oded Goldreich, and Hugo Krawczyk. Stateless Evaluation of Pseudorandom Functions: Security beyond the Birthday Barrier. In Michael J. Wiener, editor, *Advances in Cryptology - CRYPTO '99*, volume 1666 of *LNCS*, pages 270–287. Springer, 1999.

[BGM04] Mihir Bellare, Oded Goldreich, and Anton Mityagin. The Power of Verification Queries in Message Authentication and Authenticated Encryption. IACR Cryptology ePrint Archive, Report 2004/309, 2004. Available at http://eprint.iacr.org/2004/309.

[BHK+99] John Black, Shai Halevi, Hugo Krawczyk, Ted Krovetz, and Phillip Rogaway. UMAC: Fast and Secure Message Authentication. In Michael J. Wiener, editor, *Advances in Cryptology - CRYPTO '99*, volume 1666 of *LNCS*, pages 216–233. Springer, 1999.

[BI99] Mihir Bellare and Russell Impagliazzo. A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. IACR Cryptology ePrint Archive, Report 1999/024, 1999. Available at http://eprint.iacr.org/1999/024.

[BKR98] Mihir Bellare, Ted Krovetz, and Phillip Rogaway. Luby-Rackoff Backwards: Increasing Security by Making Block Ciphers Non-invertible. In Kaisa Nyberg, editor, *Advances in Cryptology - EUROCRYPT '98*, volume 1403 of *LNCS*, pages 266–280. Springer, 1998.

[BKR00] Mihir Bellare, Joe Kilian, and Phillip Rogaway. The Security of the Cipher Block Chaining Message Authentication Code. *Journal of Computer and System Sciences*, 61(3):362–399, 2000.

[BPR05] Mihir Bellare, Krzysztof Pietrzak, and Phillip Rogaway. Improved Security Analyses for CBC MACs. In Victor Shoup, editor, *Advances in Cryptology - CRYPTO 2005*, volume 3621 of *LNCS*, pages 527–545. Springer, 2005.

[BR05] John Black and Phillip Rogaway. CBC MACs for Arbitrary-Length Messages: The Three-Key Constructions. *J. Cryptology*, 18(2):111–131, 2005.

[BR06] Mihir Bellare and Phillip Rogaway. The Security of Triple Encryption and a Framework for Code-Based Game-Playing Proofs. In Serge Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 409–426. Springer, 2006. Full version available at http://eprint.iacr.org/2004/331.

[Bra82] Gilles Brassard. On Computationally Secure Authentication Tags Requiring Short Secret Shared Keys. In David Chaum, Ronald L. Rivest, and Alan T. Sherman, editors, *Advances in Cryptology - CRYPTO '82*, pages 79–86. Plenum Press, New York, 1982.

[CLL$^+$14] Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the Two-Round Even-Mansour Cipher. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology - CRYPTO 2014 (Proceedings, Part I)*, volume 8616 of *LNCS*, pages 39–56. Springer, 2014. Full version available at http://eprint.iacr.org/2014/443.

[CLP14] Benoît Cogliati, Rodolphe Lampe, and Jacques Patarin. The Indistinguishability of the XOR of k Permutations. In Carlos Cid and Christian Rechberger, editors, *Fast Software Encryption - FSE 2014*, volume 8540 of *LNCS*, pages 285–302. Springer, 2014.

[CS14] Shan Chen and John Steinberger. Tight Security Bounds for Key-Alternating Ciphers. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 327–350. Springer, 2014. Full version available at http://eprint.iacr.org/2013/222.

[DS11] Yevgeniy Dodis and John P. Steinberger. Domain Extension for MACs Beyond the Birthday Barrier. In Kenneth G. Paterson, editor, *Advances in Cryptology - EUROCRYPT 2011*, volume 6632 of *LNCS*, pages 323–342. Springer, 2011.

[Fer05] Niels Ferguson. Authentication Weaknesses in GCM. Comments submitted to NIST Modes of Operation Process, 2005. Available at http://csrc.nist.gov/groups/ST/toolkit/BCM/documents/comments/CWC-GCM/Ferguson2.pdf.

[GMS74] Edgar N. Gilbert, F. Jessie MacWilliams, and Neil J. A. Sloane. Codes which detect deception. *Bell System Technical Journal*, 53(3):405–424, 1974.

[HK97] Shai Halevi and Hugo Krawczyk. MMH: Software Message Authentication in the Gbit/Second Rates. In Eli Biham, editor, *Fast Software Encryption - FSE '97*, volume 1267 of *LNCS*, pages 172–189. Springer, 1997.

[HP08] Helena Handschuh and Bart Preneel. Key-Recovery Attacks on Universal Hash Function Based MAC Algorithms. In David Wagner, editor, *Advances in Cryptology - CRYPTO 2008*, volume 5157 of *LNCS*, pages 144–161. Springer, 2008.

[HWKS98] Chris Hall, David Wagner, John Kelsey, and Bruce Schneier. Building PRFs from PRPs. In Hugo Krawczyk, editor, *Advances in Cryptology - CRYPTO '98*, volume 1462 of *LNCS*, pages 370–389. Springer, 1998.

[JJV02] Éliane Jaulmes, Antoine Joux, and Frédéric Valette. On the Security of Randomized CBC-MAC Beyond the Birthday Paradox Limit: A New Construction. In Joan Daemen and Vincent Rijmen, editors, *Fast Software Encryption - FSE 2002*, volume 2365 of *LNCS*, pages 237–251. Springer, 2002.

[JL04] Éliane Jaulmes and Reynald Lercier. FRMAC, a Fast Randomized Message Authentication Code. 2004. Available at http://eprint.iacr.org/2004/166.

[Jou06] Antoine Joux. Authentication Failures in NIST Version of GCM. Comments submitted to NIST Modes of Operation Process, 2006. Available at http://csrc.nist.gov/groups/ST/toolkit/BCM/documents/comments/800-38_Series-Drafts/GCM/Joux_comments.pdf.

[KR00] Ted Krovetz and Phillip Rogaway. Fast Universal Hashing with Small Keys and No Preprocessing: The PolyR Construction. In Dongho Won, editor, *Information Security and Cryptology - ICISC 2000*, volume 2015 of *LNCS*, pages 73–89. Springer, 2000.

[Kra94] Hugo Krawczyk. LFSR-based Hashing and Authentication. In Yvo Desmedt, editor, *Advances in Cryptology - CRYPTO '94*, volume 839 of *LNCS*, pages 129–139. Springer, 1994.

[KVW04] Tadayoshi Kohno, John Viega, and Doug Whiting. CWC: A High-Performance Conventional Authenticated Encryption Mode. In Bimal K. Roy and Willi Meier, editors, *Fast Software Encryption - FSE 2004*, volume 3017 of *LNCS*, pages 408–426. Springer, 2004.

[Luc00] Stefan Lucks. The Sum of PRPs Is a Secure PRF. In Bart Preneel, editor, *Advances in Cryptology - EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 470–484. Springer, 2000.

[Min10] Kazuhiko Minematsu. How to Thwart Birthday Attacks against MACs via Small Randomness. In Seokhie Hong and Tetsu Iwata, editors, *Fast Software Encryption - FSE 2010*, volume 6147 of *LNCS*, pages 230–249. Springer, 2010.

[MV04] David A. McGrew and John Viega. The Security and Performance of the Galois/Counter Mode (GCM) of Operation. In Anne Canteaut and Kapalee Viswanathan, editors, *Progress in Cryptology - INDOCRYPT 2004*, volume 3348 of *LNCS*, pages 343–355. Springer, 2004.

[Pat90] Jacques Patarin. Pseudorandom Permutations Based on the DES Scheme. In Gérard D. Cohen and Pascale Charpin, editors, *EUROCODE '90*, volume 514 of *LNCS*, pages 193–204. Springer, 1990.

[Pat91] Jacques Patarin. New Results on Pseudorandom Permutation Generators Based on the DES Scheme. In Joan Feigenbaum, editor, *Advances in Cryptology - CRYPTO '91*, volume 576 of *LNCS*, pages 301–312. Springer, 1991.

[Pat08a] Jacques Patarin. A Proof of Security in $O(2^n)$ for the Xor of Two Random Permutations. In Reihaneh Safavi-Naini, editor, *Information Theoretic Security - ICITS 2008*, volume 5155 of *LNCS*, pages 232–248. Springer, 2008. Full version available at http://eprint.iacr.org/2008/010.

[Pat08b] Jacques Patarin. The "Coefficients H" Technique. In Roberto Maria Avanzi, Liam Keliher, and Francesco Sica, editors, *Selected Areas in Cryptography - SAC 2008*, volume 5381 of *LNCS*, pages 328–345. Springer, 2008.

[Pat13] Jacques Patarin. Security in $O(2^n)$ for the Xor of Two Random Permutations: Proof with the Standard H Technique. IACR Cryptology ePrint Archive, Report 2013/368, 2013. Available at http://eprint.iacr.org/2013/368.

[PC15] Gordon Procter and Carlos Cid. On Weak Keys and Forgery Attacks Against Polynomial-Based MAC Schemes. *J. Cryptology*, 28(4):769–795, 2015. Earlier version at FSE 2013.

[Rog95] Phillip Rogaway. Bucket Hashing and its Application to Fast Message Authentication. In Don Coppersmith, editor, *Advances in Cryptology - CRYPTO '95*, volume 963 of *LNCS*, pages 29–42. Springer, 1995.

[Saa12] Markku-Juhani O. Saarinen. Cycling Attacks on GCM, GHASH and Other Polynomial MACs and Hashes. In Anne Canteaut, editor, *Fast Software Encryption - FSE 2012*, volume 7549 of *LNCS*, pages 216–225. Springer, 2012.

[Sho96] Victor Shoup. On Fast and Provably Secure Message Authentication Based on Universal Hashing. In Neal Koblitz, editor, *Advances in Cryptology - CRYPTO '96*, volume 1109 of *LNCS*, pages 313–328. Springer, 1996.

[WC81] 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.

[Yas10] Kan Yasuda. The Sum of CBC MACs Is a Secure PRF. In Josef Pieprzyk, editor, *Topics in Cryptology - CT-RSA 2010*, volume 5985 of *LNCS*, pages 366–381. Springer, 2010.

[Yas11] Kan Yasuda. A New Variant of PMAC: Beyond the Birthday Bound. In Phillip Rogaway, editor, *Advances in Cryptology - CRYPTO 2011*, volume 6841 of *LNCS*, pages 596–609. Springer, 2011.

[ZWSW12] Liting Zhang, Wenling Wu, Han Sui, and Peng Wang. 3kf9: Enhancing 3GPP-MAC beyond the Birthday Bound. In Xiaoyun Wang and Kazue Sako, editors, *Advances in Cryptology - ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 296–312. Springer, 2012.

A The Encrypted Davies-Meyer Construction

In this section, we consider the Encrypted Davies-Meyer construction

$$\mathsf{EDM}[P, P'](x) = P'(P(x) \oplus x),$$

where P and P' are independent random permutations of $\{0,1\}^n$, and prove that it is secure up to roughly $2^{2n/3}$ adversarial queries. More precisely, one has the following theorem.

Theorem 5. *Let* A *be an adversary with oracle access to a function from* $\{0,1\}^n$ *to* $\{0,1\}^n$*, making at most* q *oracle queries, and returning a single bit. Then its advantage in distinguishing the EDM construction from a uniformly random function, defined as*

$$\left|\Pr\left[(P, P') \leftarrow_\$ \mathsf{Perm}(n)^2 : \mathsf{A}^{\mathsf{EDM}[P,P']} = 1\right] - \Pr\left[R \leftarrow_\$ \mathsf{Func}(n) : \mathsf{A}^R = 1\right]\right|,$$

is less than $5q^{3/2}/2^n$.

Proof. The proof uses the H-coefficients technique: the real world corresponds to $\mathsf{EDM}[P, P']$, while the ideal world corresponds to R. Fix an adversary A, and consider the transcript $\tau = ((x_1, y_1), \ldots, (x_q, y_q))$ of the queries x_i of the adversary and corresponding answers y_i (in the following, we refer to a pair (x_i, y_i) as a query). We say that a transcript is attainable if there exists a function $R \in \mathsf{Func}(n)$ such that A interacting with R results in transcript τ. We denote Θ the set of attainable transcripts. We also denote X_{re}, resp. X_{id}, the probability distribution of the transcript τ induced by the real world, resp. the ideal world.

We say that a query $(x_i, y_i) \in \tau$ is *collisioning* if $y_i = y_j$ for some $j \neq i$, otherwise we say it is *non-collisioning*. We say that an attainable transcript τ is bad if the number of collisioning queries is more than $\sqrt{q}$. Otherwise, we say that τ is good. We denote Θ_{bad}, resp. Θ_{good}, the set of bad, resp. good transcripts.

We first upper bound the probability to obtain a bad transcript in the ideal world. Since in that case the y_i's are uniformly random and independent, the expected number of collisioning queries is less than $q^2/2^n$. Hence, by Markov's inequality,

$$\Pr\left[X_{\text{id}} \in \Theta_{\text{bad}}\right] \leq \frac{q^{3/2}}{2^n}. \tag{22}$$

Consider now a good transcript τ. We need to lower bound the probability to obtain τ in the real world. We reorder the transcript as follows. Assume that the number of distinct y-coordinates in the transcript is r. Then we rewrite the transcript so that all pairs with the same y-coordinate are consecutive. The transcript is now

$$\begin{aligned}\tau = \big((&x_{1,1}, y_1), \ldots, (x_{1,q_1}, y_1),\\ &(x_{2,1}, y_2), \ldots, (x_{2,q_2}, y_2),\\ &\ldots,\\ &(x_{r,1}, y_r), \ldots, (x_{r,q_r}, y_r)\big),\end{aligned}$$

where $y_1, \ldots, y_r$ are distinct and $\sum_{i=1}^{r} q_i = q$.

In order to lower bound the probability of τ in the real world, we need to lower bound the number of pairs of permutations (P, P') such that

$$\forall i \in \{1, \ldots, r\},\ \forall j \in \{1, \ldots, q_i\},\ P'(P(x_{i,j}) \oplus x_{i,j}) = y_i.$$

For this, we will consider all possible "internal" values $z_i = (P')^{-1}(y_i)$. We say that a tuple $\mathbf{z} = (z_1, \ldots, z_r)$ of distinct values is *good* if all values $z_i \oplus x_{i,j}$ for $i \in \{1, \ldots, r\}$ and $j \in \{1, \ldots, q_i\}$ are distinct. Given a good tuple $\mathbf{z}$, the probability that

$$\begin{cases} \forall i \in \{1, \ldots, r\},\ \forall j \in \{1, \ldots, q_i\},\ P(x_{i,j}) = z_i \oplus x_{i,j}, \\ \forall i \in \{1, \ldots, r\},\ P'(z_i) = y_i \end{cases}$$

is exactly

$$\frac{1}{(2^n)_q (2^n)_r}. \tag{23}$$

(This is simply the probability that P satisfies $q_1 + \ldots + q_r = q$ equations and P' satisfies r equations.)

We can lower bound the number of good tuples $\mathbf{z}$ as follows:

- there are at least 2^n possibilities for z_1;
- once z_1 is fixed, there are at least $2^n - 1 - q_1 q_2$ possibilities for z_2, since z_2 must be different from z_1 and from $z_1 \oplus x_{1,j} \oplus x_{2,j'}$ for all $j \in \{1, \ldots, q_1\}$ and all $j' \in \{1, \ldots, q_2\}$;

– once z_1 and z_2 are fixed, there are at least $2^n - 2 - (q_1 + q_2)q_3$ possibilities for z_3, since z_3 must be different from z_1, z_2, $z_1 \oplus x_{1,j} \oplus x_{3,j'}$ for all $j \in \{1, \ldots, q_1\}$ and all $j' \in \{1, \ldots, q_3\}$, and from $z_2 \oplus x_{2,j} \oplus x_{3,j'}$ for all $j \in \{1, \ldots, q_2\}$ and all $j' \in \{1, \ldots, q_3\}$;
– etc.

Hence, the number of good tuples $\mathbf{z}$ is at least

$$\prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{j=1}^{i} q_j \right) .$$

Hence, summing probability (23) over all possibilities for $\mathbf{z}$, the probability to get the transcript in the real world satisfies

$$\Pr\left[X_{\mathrm{re}} = \tau\right] \geq \frac{\prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{j=1}^{i} q_j\right)}{(2^n)_q (2^n)_r}.$$

Since the probability to obtain τ in the ideal world is simply $1/(2^n)^q$, the ratio of probabilities is at least

$$\begin{aligned}
\rho \stackrel{\text{def}}{=} \frac{\Pr\left[X_{\mathrm{re}} = \tau\right]}{\Pr\left[X_{\mathrm{id}} = \tau\right]} &\geq \frac{(2^n)^q \prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{j=1}^{i} q_j\right)}{(2^n)_q (2^n)_r} \\
&= \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i}\right) \prod_{i=0}^{r-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i}\right) .
\end{aligned}$$

In order to further lower bound this ratio ρ, we need to distinguish collisioning queries from non-collisioning ones. Up to reordering the transcript, we assume that non-collisioning queries come first, and we let $s \in \{0, \ldots, r\}$ be the integer such that $q_i = 1$ for $i \in \{1, \ldots, s\}$, and $q_i > 1$ for $i \in \{s+1, \ldots, r\}$. Note that since the transcript is good,

$$\sum_{i=s+1}^{r} q_i \leq \sqrt{q}. \tag{24}$$

Then

$$\begin{aligned}
\rho &\geq \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i}\right) \prod_{i=0}^{s-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i}\right) \\
&= \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i}\right) \prod_{i=0}^{s-1} \left(1 - \frac{i}{2^n - i}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i}\right) \\
&\geq \prod_{i=0}^{q-1} \left(1 - \frac{i^2}{(2^n - i)^2}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} q}{2^n - i}\right) \\
&\geq \prod_{i=0}^{q-1} \left(1 - \frac{i^2}{(2^n - q)^2}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} q}{2^n - q}\right)
\end{aligned}$$

$$\geq \left(1 - \frac{q^3}{3(2^n - q)^2}\right)\left(1 - \frac{q\sum_{i=s+1}^{r} q_i}{2^n - q}\right)$$
$$\geq \left(1 - \frac{4q^3}{3 \cdot 2^{2n}}\right)\left(1 - \frac{2q^{3/2}}{2^n}\right),$$

where for the last inequality we used $q \leq 2^n/2$ and (24). Since $q^3/2^{2n} \leq q^{3/2}/2^n$ by our assumption that $q^{3/2} \leq 2^n/4$, we obtain

$$\frac{\Pr\left[X_{\mathrm{re}} = \tau\right]}{\Pr\left[X_{\mathrm{id}} = \tau\right]} \geq 1 - \frac{4q^{3/2}}{2^n}. \tag{25}$$

Combining (22) and (25) with Lemma 1, we obtain that the distinguishing advantage is at most $5q^{3/2}/2^n$, as announced. □