How to Thwart Birthday Attacks against MACs via Small Randomness

Kazuhiko Minematsu

NEC Corporation, 1753 Shimonumabe, Nakahara-Ku, Kawasaki, Japan,
k-minematsu@ah.jp.nec.com

Abstract. The security of randomized message authentication code, MAC for short, is typically depending on the uniqueness of random initial vectors (IVs). Thus its security bound usually contains $O(q^2/2^n)$, when random IV is n bits and q is the number of MACed messages. In this paper, we present how to break this birthday barrier without increasing the randomness. Our proposal is almost as efficient as the well-known Carter-Wegman MAC, uses n-bit random IVs, and provides the security bound roughly $O(q^3/2^{2n})$. We also provide blockcipher-based instantiations of our proposal. They are almost as efficient as CBC-MAC and the security is solely based on the pseudorandomness of the blockcipher.

Key words: Message Authentication Code, Birthday Bound, Mode of Operation

1 Introduction

Message Authentication Code. Message Authentication Codes (MACs) are symmetric cryptographic functions used to ensure the authenticities of messages. Its usage is as follows. When Alice wants to send a message M, she computes a MAC function that accepts M and a secret key, K, and possibly an auxiliary variable called IV (stands for initial vector), and obtains an authentication tag T as an output. Then she sends (IV, M, T) to Bob, who shares K. Bob verifies if (IV, M, T) is authentic or not by computing the MAC using (IV, M) and K to obtain the local tag T', and see if T' matches T. If IV is a nonce, e.g., a counter, the MAC is said to be stateful. If IV is random, the MAC is said to be (stateless but) randomized. An adversary observes valid (IV, M, T) tuples and tries to make a forgery, i.e., a new tuple (IV', M', T') which is determined as authentic by Bob. If this is hard, we say the MAC is strongly unforgeable [2].

Security of Hash-then-Mask. To build an IV-based MAC, a common approach is Carter and Wegman's one [11]: it uses an ϵ-almost XOR universal (ϵ-AXU, see Sect. 2) hash function $H : \{0,1\}^* \to \{0,1\}^\pi$, and a pseudorandom function (PRF) $F : \{0,1\}^n \to \{0,1\}^\pi$. It produces a π-bit tag $T = F(IV) \oplus H(M)$ for message M using n-bit IV. We call this structure Hash-then-Mask (HtM). It is denoted by $\Pi^{\mathrm{rnd}}_{n,\pi,\epsilon}$ when IV is random, and $\Pi^{\mathrm{ctr}}_{n,\pi,\epsilon}$ when IV is a nonce.

Let us take a close look at the security of HtM against attacks with q tagging queries and q_v verification queries (see Sect. 2), where the goal of attack is to

break the strong unforgeability. For $\Pi^{\mathrm{ctr}}_{n,\pi,\epsilon}$, it is well known that the probability of a forgery is at most $q_v\epsilon$ for any $q \leq 2^n$ [4][9], except a term for the computational security of F. However, in case of $\Pi^{\mathrm{rnd}}_{n,\pi,\epsilon}$ the probability of forgery is degraded to $q^2/2^n + q_v\epsilon$ as IVs may collide with probability about $q^2/2^n$, that is, the birthday bound[1]. In fact, it is easy to prove that the above bound is tight for q (see Sect. 3). This degradation is non-negligible when n is relatively small, say 64. In addition, as pointed out by many researchers [3][16] the use of nonce is sometimes impractical. Hence it is natural to ask if we could break the above-mentioned birthday bound without being stateful. A trivial solution is to use a longer random IV. The randomized HtM with $2n$-bit IV ($\Pi^{\mathrm{rnd}}_{2n,\pi,\epsilon}$) provides the bound $q^2/2^{2n} + q_v\epsilon$, where F is a PRF with $2n$-bit inputs. However, this is problematic since (1) long random IV forces increased communication cost and sender's effort for generating randomness, and (2) the need for $2n$-bit-input PRF instead of n-bit-input one limits the applicability.

The second problem can be avoided by using MACRX$_3$ [3]. It uses three n-bit-input PRFs and an ϵ-AXU hash of π-bit output, and achieves $O(q^3/2^{3n} + q_v\epsilon)$-security[2]. Unfortunately, MACRX$_3$ requires an even longer, $3n$-bit random IV. Thus it still fails to avoid the first problem. As solutions to the both problems, RMAC [16] and FRMAC [17] are known. They use an n-bit random IV and an n-bit blockcipher. The bound of RMAC is $O(\sigma/2^n)$ where σ is the total message blocks for all tagging and verification queries. FRMAC has a similar bound. However, their security proofs are based on a controversial assumption on the internal block cipher [29][18].

Our Contribution. From above discussion, what is important is to build a randomized MAC with n-bit IV and has security bound better than $O(q^2/2^n)$ based on the standard assumptions. For this purpose, we first allow us to use $2n$-bit-input PRF, combined with a universal hash having n-bit output. Our proposal, called RWMAC, is just a randomized version of a nonce-based MAC called WMAC [8] (and almost the same as a function appeared in the proof of FRMAC [17]). With n-bit random IV and π-bit tag, RWMAC has $O(\epsilon q^2/2^n + q_v(\epsilon + 1/2^\pi))$-security when the universal hash is ϵ-Almost universal (ϵ-AU). As $\epsilon \geq 1/2^n$, we can achieve $O(q^2/2^{2n} + q_v/2^n)$-security at best. Although our proposal itself is not so new, we think our security proof is new and non-trivial.

Naturally, the next step is to build a randomized MAC with n-bit random IV and n-bit-input PRF, which appears much more challenging. We present a solution, called Enhanced Hash-then-Mask (EHtM), which will be the main contribution of this paper. EHtM is *very* efficient, as it uses only two calls of n-bit-input PRFs and one call of an ϵ-AXU hash with n-bit output. Tag length π can be set to any value up to the output length of PRF. In return for this excellent property, the security bound is $O(\epsilon q^3/2^n + q_v(\epsilon + 2^{-\pi}))$, thus $O(q^3/2^{2n} + q_v/2^n)$ at best (when $\epsilon = O(1/2^n)$ and $\pi = n$). Hence, our scheme certainly provides a

[1] "Birthday bound" is somewhat confusing since randomized MAC has many parameters, such as tag length, IV length, etc. In this paper, we exclusively use this word to express the term $O(q^2/2^n)$ in the bound of randomized MACs with n-bit IVs.

[2] In the sense of weak unforgeability. See Sect.2 for definition.

security beyond the birthday bound, however, its bound is generally inferior to that of RWMAC.

The profiles of randomized MACs[3] are briefly summarized in Table 1. Table 1 clearly shows that the complexity (both computation and communication) of EHtM is the closest to that of the original randomized HtM among others.

Mode of Operation. EHtM is a generic construction. This generality allows us to various instantiations. Among them, we present two blockcipher modes called MAC-R1 and MAC-R2. Their complexities are almost the same as that of CBC-MAC. To prove its security, we only require that the underlying blockcipher is a pseudorandom permutation (PRP). This is a crucial difference from RMAC, which is also based on CBC-MAC but requires the ideal-cipher model for its security, which is highly problematic as shown by, e.g., Knudsen and Kohno [18]. The concrete bounds of MAC-R1 and MAC-R2 are slight worse than the original EHtM using n-bit PRFs. Still, there is a remarkable gain from CBC-MAC and its variants. We think our proposals will be good practical MACs using 64-bit blockciphers, thus suited to resource-constrained environments. A detailed, quantitative comparison will be given in Sect. 6.3.

Table 1. Profiles of randomized MACs. We set $\pi = n$ for the compatibility with RMAC and FRMAC. We assume ℓn-bit messages. $\mathrm{H_u}[i,j]$ ($\mathrm{H_{xu}}[i,j]$) denotes ϵ-AU (ϵ-AXU) hash function of i-bit input and j-bit output. $\mathrm{F}[i,j]$ denotes a PRF of i-bit input and j-bit output, and $\mathrm{P}[i]$ denotes an i-bit keyed permutation, i.e., a blockcipher. In deriving the bounds of RWMAC and RMAC, we use $\sigma \le \ell(q+q_v)$ for simplicity. The symbol ▲ indicates that the security proof requires a stronger assumption than the PRP, such as the ideal-cipher model, for $\mathrm{P}[n]$.

MAC	Rand	Efficiency	Security
Randomized Hash-then-Mask	n	$1\mathrm{H_{xu}}[\ell n, n] + 1\mathrm{F}[n,n]$	$O(q^2/2^n + q_v\epsilon)$
$\mathrm{MACRX_3}$ [3]	$3n$	$1\mathrm{H_{xu}}[\ell n, n] + 3\mathrm{F}[n,n]$	$O(q^3/2^{3n} + q_v\epsilon)$
RMAC [16]	n	$(\ell+1)\mathrm{P}[n]$	▲$O(\ell(q+q_v)/2^n)$
FRMAC [17]	n	$1\mathrm{H_u}[\ell n, n] + 1\mathrm{P}[n]$	▲$O(\ell(q+q_v)\epsilon)$
RWMAC (this paper, similar to [8])	n	$1\mathrm{H_u}[\ell n, n] + 1\mathrm{F}[2n,n]$	$O(\epsilon q^2/2^n + q_v\epsilon)$
Enhanced Hash-then-Mask (this paper)	n	$1\mathrm{H_{xu}}[\ell n, n] + 2\mathrm{F}[n,n]$	$O(\epsilon q^3/2^n + q_v\epsilon)$

2 Preliminaries

Basic Notations. A random variable and its sampled value are written by a capital and the corresponding small letters. A sequence of random variables is

[3] A randomized MAC of Dodis et al. [12] also aims at reducing the bound via small randomness. However the scope is different from us. Their purpose is to reduce the security degradation with respect to ℓ (not q) due to the use of non-optimal universal hash. Their proposal still contains $O(q^2/2^n)$ if n-bit universal hash is used.

written as $X^i \stackrel{\text{def}}{=} (X_1, X_2, \ldots, X_i)$. $\{0,1\}^n$ is denoted by Σ^n, and Σ^* denotes the set of all finite-length bit sequences, including the empty string ϕ (which is a unique element of Σ^0). The bit length of x is denoted by $|x|$, with $|\phi| = 0$. A concatenation of two binary sequences, x and y, is written as $x\|y$. For any x and $\pi \leq |x|$, $\mathtt{chop}_\pi(x)$ is the first π bits of x.

A keyed function is written by a capital letter, and if it has n-bit inputs and m-bit outputs it is written as $F : \Sigma^n \to \Sigma^m$, i.e., we omit the description of key space. $F(*\|w)$ is a keyed function $\Sigma^{n-|w|} \to \Sigma^m$. In particular, the uniform random function (URF) : $\Sigma^n \to \Sigma^m$ is denoted by $\mathsf{R}_{n,m}$. This is a random function whose distribution is uniform over $\{f : \Sigma^n \to \Sigma^m\}$. The n-bit uniform random permutation (URP), denoted by P_n, is a random permutation with a uniform distribution over all permutations of Σ^n.

Definition 1. *Let $H : \Sigma^* \to \Sigma^n$ be a keyed function. If $\Pr[H(x) = H(x')] \leq \epsilon(\ell)$ holds for any distinct x, x' with $\max\{|x|, |x'|\} \leq \ell n$, where probability is defined by H's key, H is said to be $\epsilon(\ell)$-almost universal ($\epsilon(\ell)$-AU). In addition, if $\Pr[H(x) \oplus H(x') = y] \leq \epsilon(\ell)$ holds for any $y \in \Sigma^n$ and distinct x, x' with $\max\{|x|, |x'|\} \leq \ell n$, H is said to be an $\epsilon(\ell)$-almost XOR universal ($\epsilon(\ell)$-AXU). We also say H is universal (XOR-universal) if $\epsilon(\ell)$ is minimum, i.e., when H is $1/2^n$-AU ($1/2^n$-AXU).*

For any keyed function F, $\mathtt{Adv}^{\mathtt{prf}}_F(q, \tau)$ denotes the maximum advantage [1] in distinguishing F from a URF having the same input/output domains using q chosen-plaintext queries and computational complexity τ. Moreover, for any keyed permutation E over Σ^n, $\mathtt{Adv}^{\mathtt{prp}}_E(q, \tau)$ denotes the maximum advantage in distinguishing E from P_n.

Definition 2. *A randomized MAC function with η-bit randomness and π-bit tag is defined as a keyed function $\mathbf{F} : \Sigma^\eta \times \Sigma^* \to \Sigma^\pi$. A query to the tagging oracle (called a tagging query) is a message $M \in \Sigma^*$, and the corresponding answer is $(U, T) \in \Sigma^\eta \times \Sigma^\pi$, where U is independent and uniform over η bits, and $T = \mathbf{F}(U, M)$. A query to the verification oracle (called a verification query) is a tuple $(\widetilde{U}, \widetilde{M}, \widetilde{T})$ and the corresponding answer, written as a binary digit B, is 1 if $\widetilde{T} = \mathbf{F}(\widetilde{U}, \widetilde{M})$ and 0 otherwise.*

Here, $\mathbf{F}$ does not produce U on its own. For any $\mathbf{F}$ we implicitly assume the uniform distribution of U. In a verification query, $\widetilde{U}$ can be arbitrarily chosen. As mentioned in Introduction, the adversary's goal is to create a forgery in the sense of strong unforgeability [2] defined as follows.

Definition 3. *A (q, q_v, ℓ, τ)-forger, $\mathcal{A}$, against a randomized MAC, $\mathbf{F} : \Sigma^\eta \times \Sigma^* \to \Sigma^\pi$, is an entity that performs q tagging queries and q_v verification queries, where every message is at most ℓn-bit and $\mathcal{A}$'s total computational complexity is τ. We use subscripts to express the ordinal number of queries, e.g., M_i denotes the i-th tagging query. If $(\widetilde{U}_j, \widetilde{M}_j, \widetilde{T}_j) \neq (U_i, M_i, T_i)$, $i = 1, \ldots, q$, and $B_j = 1$ holds for some $j \in \{1, \ldots, q_v\}$, $(\widetilde{U}_j, \widetilde{M}_j, \widetilde{T}_j)$ is called a successful forgery.*

Note that M_i can depend on U^{i-1}, M^{i-1}, and T^{i-1} but not depend on U_i.
Strong and Weak Unforgeabilities. If we require a stricter condition that $\widetilde{M}_j \neq M_i$ for $i = 1, \ldots, q$, we call the corresponding security notion the weak unforgeability. This notion is defined as (mere) unforgeability by Bellare et al. [2]. See [2] for the technical differences in strong and weak unforgealibities.

Definition 4. *For any forger $\mathcal{A}$ and randomized MAC $\mathbf{F}$, the forgery probability is the probability that $\mathcal{A}$ produces at least one successful forgery (in the sense of Def. 3) for $\mathbf{F}$. The maximum forgery probability for all (q, q_v, ℓ, τ)-forgers is denoted by $\mathrm{FP}_{\mathbf{F}}(q, q_v, \ell, \tau)$. By omitting τ we mean the maximum information-theoretic forgery probability, i.e., $\mathrm{FP}_{\mathbf{F}}(q, q_v, \ell)$ means $\mathrm{FP}_{\mathbf{F}}(q, q_v, \ell, \infty)$.*

As pointed out by [6], if we focus on the first successful forgery, we only need to consider forgers that first perform q tagging queries and then perform q_v verification queries. I.e., the game is divided into the consecutive two phases; the tagging and verification phases. This restriction does not increase the chance of single successful forgery. Also, the verification phase can be defined as a batch process, i.e., $(\widetilde{U}_j, \widetilde{M}_j, \widetilde{T}_j)$ is a (possibly non-deterministic) function of (U^q, M^q, T^q) and not dependent on $(\widetilde{U}^{j-1}, \widetilde{M}^{j-1}, \widetilde{T}^{j-1}, B^{j-1})$. However, these conventions will not work if we focus on other security notions, see [8][23].

3 Randomized WMAC

Limitation of Hash-then-Mask. Let us consider a randomized HtM with n-bit IV, π-bit tag, defined as $\Pi^{\mathrm{rnd}}_{n,\pi,\epsilon}$ in Introduction. The components are $H : \Sigma^* \to \Sigma^\pi$ which is ϵ-AXU and $F : \Sigma^n \to \Sigma^\pi$ which is URF. Then we have

$$\mathrm{FP}_{\Pi^{\mathrm{rnd}}_{n,\pi,\epsilon(\ell)}}(q, q_v, \ell) \leq q^2/2^{n+1} + \epsilon(\ell) q_v, \tag{1}$$

since the bound of $\Pi^{\mathrm{ctr}}_{n,\pi,\epsilon}$ is ϵq_v [9] and the forgery probability under random IVs is at most the sum of forgery probability under *distinct* random IVs (i.e., nonce) and the probability of IV collision, which is at most $\binom{q}{2}/2^n \leq q^2/2^{n+1}$. In fact, the above bound is tight as $2^{n/2}$ tagging queries are enough to break $\Pi^{\mathrm{rnd}}_{n,\pi,\epsilon}$. The attack is as follows:

1. Make j tagging queries with distinct M^j where a collision $U_i = U_j$ for some $i < j$ occurs.
2. Let $M_{j+1} = M_j$. Check if $U_{j+1} \neq U_j$ holds (otherwise try another query with the same message).
3. Make a verification query as $(\widetilde{U}, \widetilde{M}, \widetilde{T}) = (U_{j+1}, M_i, T_i \oplus T_j \oplus T_{j+1})$.

As $T_i \oplus T_j \oplus T_{j+1} = F(U_{j+1}) \oplus H(M_i)$, $\widetilde{T}$ is a valid tag for a new tuple (U_{j+1}, M_i)[4]. The attack succeeds with probability almost 1 if we use $2^{n/2}$ queries in the step

[4] Here we break the strong unforgeability: it is open if the bound is also tight for the weak unforgeability.

1. Since the attack does not exploit any specific properties of H and F, it works for any randomized HtM[5]. Hence, to break the bound $O(q^2/2^n)$ while keeping the n-bit random IV, we need a different structure from Hash-then-Mask.

Randomized WMAC. To avoid the above attack, a promising solution is to process the n-bit hash value, $S = H(M)$, and the n-bit random IV, U, together with a $2n$-bit-input PRF, G. More precisely, the tag $T \in \Sigma^\tau$ for M is generated as $T = G(U, H(M))$, where $H : \Sigma^* \to \Sigma^n$ is $\epsilon(\ell)$-AU and $G : \Sigma^{2n} \to \Sigma^\tau$ is a PRF. This MAC is denoted by RWMAC$[H, G]$ as it is a randomized version of WMAC [8], a nonce-based MAC. Indeed, RWMAC offers a very high security, since neither an S-collision nor a U-collision can be noticed by adversary, unless both collisions occur simultaneously. The security bound is as follows[6].

Theorem 1. *If H is $\epsilon(\ell)$-AU and $q \le \min\{2^{n-2}, \sqrt{2^n \cdot \epsilon(\ell)^{-1}}\}$,*

$$\mathrm{FP}_{\mathrm{RWMAC}[H,G]}(q, q_v, \ell, \tau) = \mathtt{Adv}^{\mathtt{prf}}_G(q + q_v, \tau + O(q + q_v)) + q^2 \frac{\epsilon(\ell)}{2^{n+1}} + q_v \left(2(n-1)\epsilon(\ell) + \frac{1}{2^\tau}\right).$$

The proof of Theorem 1 is in Appendix A. The structure of the proof is the same as that of our main theorem (Theorem 2), but details are much simpler.

4 Enhanced Hash-then-Mask

Although RWMAC provides a very high security, a big problem still remains: it needs G, a PRF with $2n$-bit input, while the original HtM is based on a PRF with n-bit input. One may try some domain extension scheme of an n-bit-input PRF to obtain a $2n$-bit-input PRF. However, most known schemes such as CBC-MAC, are only $O(q^2/2^n)$-secure, thus can not be used for our purpose. One workable scheme of Maurer [21] is a composition of a keyed function that diffuses a $2n$-bit input to a cn-bit output for some $c \ge 2$ and an encryption function consisting of c PRFs aligned parallel. The output is the sum of each c PRFs' outputs. The security bound is $O(q^{c+1}/2^{cn})$ [21]. However, it is still cumbersome to implement this diffuse-encrypt-xor scheme, as the diffusion must be $2c$-locally-uniform [21], which is much costly than the universal hash functions even for a small c.

Nevertheless, there seems a chance of a simpler domain extension scheme, because inputs to G of RWMAC can not be arbitrarily chosen. We will prove that this intuition is true: $2n$-bit PRF of RWMAC can be safely substituted with an extremely simple function using two n-bit PRFs. The concrete proposal and its security bound is in the following.

[5] This attack has some similarities to the L-collision attack by Semanko [26], though the targets of attacks are different.

[6] An equivalent to RWMAC was appeared in Lemma 4 of [17] and the bound $O(\sigma\epsilon(\ell))$ was claimed, though we did not scrutinize the proof.

Definition 5. *Let $H : \Sigma^* \to \Sigma^n$ and $F_i : \Sigma^n \to \Sigma^n$ for $i = 1, 2$. The enhanced hash-then-mask (EHtM) with π-bit tags (for some $\pi \leq n$) is defined as* $\mathrm{EHtM}[H, F_1, F_2](U, M) \stackrel{\mathrm{def}}{=} \mathtt{chop}_\pi(U, F_1(U) \oplus F_2(H(M) \oplus U))$ *for message $M \in \Sigma^*$, where $U \in \Sigma^n$ is independent and uniformly random.*

Theorem 2. *Let $H : \Sigma^* \to \Sigma^n$ be $\epsilon(\ell)$-AXU. Let F_1 and F_2 be independently-keyed instances of $F : \Sigma^n \to \Sigma^n$. Then we have*

$$\mathrm{FP}_{\mathrm{EHtM}[H,F_1,F_2]}(q, q_v, \ell, \tau) \leq 2\mathtt{Adv}^{\mathtt{prf}}_F(q + q_v, \tau') + \frac{q^3}{6}\left(\frac{\epsilon(\ell)}{2^n} + \frac{1}{2^{3n}}\right) + q_v\left(4\epsilon(\ell) + \frac{1}{2^\pi}\right),$$

if $q \leq 3\left(\epsilon(\ell)/2^n + 1/2^{3n}\right)^{-1/3}$. Here $\tau' = \tau + O(q + q_v)$.

Hence, EHtM is secure if $q \ll (6 \cdot 2^n \cdot \epsilon(\ell))^{-1/3}$ and $q_v \ll \min\{2^\pi, \epsilon(\ell)^{-1}\}$ hold. In other words, EHtM guarantees about $2n/3$-bit security for q and π-bit security for q_v, if $\epsilon \sim 1/2^n$.

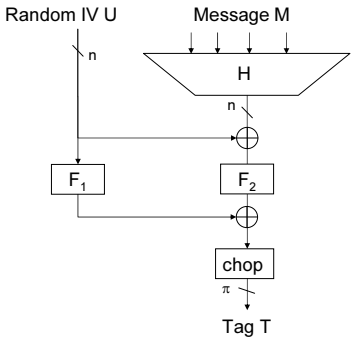

Fig. 1. Enhanced Hash-then-Mask.

5 Proof of Theorem 2

Overview. Let us denote two independent n-bit block URFs by $\mathsf{R}^{(1)}$ and $\mathsf{R}^{(2)}$. We define EH as $\mathrm{EHtM}[H, \mathsf{R}^{(1)}, \mathsf{R}^{(2)}]$ with an $\epsilon(\ell)$-AXU hash, H, and assume some $\pi \leq n$. We here prove a bound of $\mathrm{FP}_{\mathsf{EH}}(q, q_v, \ell)$. Computational counterpart is easy, thus omitted.

We first provide an intuition for the proof. Let $S_i = U_i \oplus H(M_i)$ for i-th tagging query. We observe that the finalization of EH, $(U, S) \to \mathsf{R}^{(1)}(U) \oplus \mathsf{R}^{(2)}(S)$, is indistinguishable from a $2n$-bit-input URF, if set $\mathcal{G} = \{(U_1, S_1), \ldots, (U_q, S_q)\}$ satisfies two linear conditions. These conditions are related to the linear independence of a *characteristic vector* matrix formed by $\mathcal{G}$, but weaker than that.

Here, if we use the identical URF for processing of U and S, we need the linear independence as in the proof of similar structures [3][21]. We show that, with $q \approx 2^{n/2}$ tagging queries to EH the probabilities of violating these conditions are negligible: the one is $O(q^3\epsilon(\ell)/2^n)$ and the other is $O(q^2\epsilon(\ell)/2^n)$. We also show that, if the above-mentioned conditions are satisfied for tagging phase, the forgery probability is $O(q_v(\alpha\epsilon(\ell) + 2^{-\pi}))$, where α is the size of largest class of U (i.e. there is an α-collision but not $(\alpha+1)$-collision) in the tagging phase. As Us are perfectly random, the probability of $(\alpha+1)$-collision is bounded by $O(q^{\alpha+1}2^{-n\alpha})$, thus taking $\alpha = 2$ will suffice.

Setup. Let $\mathrm{Hw}(\mathbb{V})$ denote the Hamming weight of a binary sequence $\mathbb{V}$, and let $\mathrm{Hw}(\mathbb{V}, \mathbb{V}')$ be $(\mathrm{Hw}(\mathbb{V}), \mathrm{Hw}(\mathbb{V}'))$ for a pair $(\mathbb{V}, \mathbb{V}')$. For $x \in \Sigma^n$, we use $\lambda(x) \in \Sigma^{2^n}$ to denote its characteristic vector (CV) by seeing x as an integer in $[0, \ldots, 2^n - 1]$. I.e., $\mathrm{Hw}(\lambda(x)) = 1$ and the bit 1 is in the x-th coordinate of $\lambda(x)$. For $X^q \in (\Sigma^n)^q$ and $\mathcal{I} \subseteq \{1, \ldots, q\}$, we use $\bigoplus_{\mathcal{I}} \lambda(X)$ to denote $\bigoplus_{i\in\mathcal{I}} \lambda(X_i)$. Let $\mathcal{Q} \subseteq \{1, \ldots, q\}$ denote the index set of unique (U, M) pairs, i.e., for any $i \neq j$, $i, j \in \mathcal{Q}$, $(U_i, M_i) \neq (U_j, M_j)$ holds. Note that, if $i \notin \mathcal{Q}$ there exists $j \in \mathcal{Q}$ with $(U_i, M_i, T_i) = (U_j, M_j, T_j)$, and thus all transcripts outside $\mathcal{Q}$ are useless for forgers. Here, $\mathcal{Q}$ is a random variable whose probability is defined by EH and the forger, and we assume $\mathcal{Q}$ is uniquely determined for any fixed $(U^q, M^q) = (u^q, m^q)$. We will use the following probabilistic events defined on $\{(U_i, S_i)\}_{i\in\mathcal{Q}}$, where $S_i = U_i \oplus H(M_i)$ as mentioned.

- Collision-freeness: $\mathrm{CF}_q \stackrel{\mathrm{def}}{=} [(U_i, S_i) \neq (U_j, S_j)$ for all distinct $i, j \in \mathcal{Q}]$.
- Linear independence:
  $\mathrm{LID}_q \stackrel{\mathrm{def}}{=} [\mathrm{Hw}(\bigoplus_{\mathcal{I}} \lambda(U), \bigoplus_{\mathcal{I}} \lambda(S)) \neq (0,0)$ for all $\mathcal{I} \subseteq \mathcal{Q}$, $|\mathcal{I}| =$ even $\geq 2]$.
- Non-two-vulnerability :
  $\mathrm{NTV}_q \stackrel{\mathrm{def}}{=} [\mathrm{Hw}(\bigoplus_{\mathcal{I}} \lambda(U), \bigoplus_{\mathcal{I}} \lambda(S)) \neq (1,1)$ for all $\mathcal{I} \subseteq \mathcal{Q}$, $|\mathcal{I}| =$ odd $\geq 3]$.
- The size of U's largest equivalent class is at most α:
  $\mathrm{EQS}(\alpha) \stackrel{\mathrm{def}}{=} [\max_i \mathsf{ec}(U_i) \leq \alpha]$, where $\mathsf{ec}(U_i) = |\{j \in \{1, \ldots, q\} : U_j = U_i\}|$.

For convenience, when $|\mathcal{Q}| = 1$, CF_q and LID_q are defined as true. When $|\mathcal{Q}| \leq 2$, NTV_q is defined the same as LID_q. With this convention, $\mathrm{NTV}_q \to \mathrm{LID}_q \to \mathrm{CF}_q$ holds true (proof for $|\mathcal{Q}| \leq 2$ is trivial, and proof for $|\mathcal{Q}| \geq 3$ is obtained via taking contraposition). For a forger $\mathcal{A}$ and a MAC $\mathbf{F}$, let $P^{\mathcal{A}\diamond\mathbf{F}}$ denote the probability space defined by $\mathcal{A}$ and $\mathbf{F}$ (following Defs. 2 and 3). Furthermore, we define $\nu_{q,q_v,\ell}(\mathbf{F}, \mathcal{E}) \stackrel{\mathrm{def}}{=} \max_{\mathcal{A}:(q,q_v,\ell)\text{-forger}} P^{\mathcal{A}\diamond\mathbf{F}}(\mathcal{E})$ as the maximum probability of event $\mathcal{E}$. The maximum conditional probability of $\mathcal{E}$ given another condition $\mathcal{E}'$ is similarly defined and denoted by $\nu_{q,q_v,\ell}(\mathbf{F}, \mathcal{E}|\mathcal{E}')$. We also define a weak form of adversary. If $\mathcal{A}$'s tagging and verification queries are independent of T^q, i.e., M_i is made from $U^{i-1}M^{i-1}$ and $(\widetilde{U}_j, \widetilde{M}_j, \widetilde{T}_j)$ is made from (U^q, M^q) for all $i \leq q$ and $j \leq q_v$, $\mathcal{A}$ is said to be T-independent[7]. We define $\mu_{q,\ell}(\mathbf{F}, \mathcal{E})$ as the maximum probability of $\mathcal{E}$ under all T-independent (q, q_v, ℓ)-forgers. If $\mathcal{E}$ is defined for tagging phase (that is, the probability of $\mathcal{E}$ is independent of the result

[7] Here, T-independent forger is stronger than non-adaptive one, who determines M^q independent of (U^q, T^q).

of verification phase), we simply write $\nu_{q,\ell}(\mathbf{F},\mathcal{E})$ or $\mu_{q,\ell}(\mathbf{F},\mathcal{E})$. For $i=1,\ldots,q_v$, let SUC_i denote the event $B_i=1$ (see Def. 3) and let $\mathrm{SUC} \stackrel{\mathrm{def}}{=} \mathrm{SUC}_1 \vee \cdots \vee \mathrm{SUC}_{q_v}$. Now we have

$$\begin{aligned}&\mathrm{FP}_{\mathsf{EH}}(q,q_v,\ell) = \nu_{q,q_v,\ell}(\mathsf{EH},\mathrm{SUC})\\ &\le \nu_{q,q_v,\ell}(\mathsf{EH},\mathrm{SUC}|\mathrm{EQS}(\alpha)\wedge \mathrm{NTV}_q) + \nu_{q,\ell}(\mathsf{EH},\overline{\mathrm{EQS}(\alpha)}) + \nu_{q,\ell}(\mathsf{EH},\overline{\mathrm{NTV}_q}). \quad (2)\end{aligned}$$

In the following, we analyze each of the three terms in the r.h.s. of Eq. (2).

Analysis of the Third Term. Let RW be an idealized RWMAC with n-bit IV and π-bit tag, defined as $\mathsf{RW}(U,M) = \mathsf{R}_{2n,\pi}(U, U\oplus H(M))$, where $H:\Sigma^*\to\Sigma^n$ is the same as one used by EH. NTV_q and CF_q are similarly defined with $S_i = U_i \oplus H(M_i)$.

Proposition 1. *Let* $\mathsf{Func}\in\{\mathsf{EH},\mathsf{RW}\}$. *Then for* $\mathcal{E}\in\{\mathrm{LID}_q,\mathrm{NTV}_q\}$ *we have*

$$P^{\mathsf{Func}}(T_q = t_q | U^q = u^q, M^q = m^q, T^{q-1} = t^{q-1}, \mathcal{E}) = \frac{1}{2^\pi} \quad (3)$$

holds for all possible arguments (t_q, t^{q-1}, u^q, m^q), *as long as* $(u_q,m_q)\neq(u_i,m_i)$ *for all* $i\le q-1$ *(that is,* $q\in\mathcal{Q}$*). Moreover,*

$$\nu_{q,\ell}(\mathsf{EH},\overline{\mathrm{NTV}_q}) = \nu_{q,\ell}(\mathsf{RW},\overline{\mathrm{NTV}_q}) = \mu_{q,\ell}(\mathsf{RW},\overline{\mathrm{NTV}_q}) \text{ holds.} \quad (4)$$

Proof. The proof is based on Maurer's methodology [21]. See Appendix B.

Proposition 2. $\mathrm{CF}_q\wedge\overline{\mathrm{NTV}_q}$ *is equivalent to the event that there exist distinct* $i,j,k\in\{1,\ldots,q\}$, *satisfying* $U_i=U_j\neq U_k$ *and* $S_i\neq S_j=S_k$ *with* $M_i\neq M_j\neq M_k$ *(here* $M_i=M_k$ *is possible), and does not exist distinct* $i',j'\in\{1,\ldots,q\}$ *such that* $(U_{i'},S_{i'})=(U_{j'},S_{j'})$ *with* $M_{i'}\neq M_{j'}$.

Proof. See Appendix C.

Let $\mathbb{T}$ be the set of all T-independent (q,q_v,ℓ)-forgers. Now we have

$$\begin{aligned}&\mu_{q,\ell}(\mathsf{RW},\overline{\mathrm{NTV}_q}) \le \mu_{q,\ell}(\mathsf{RW},\overline{\mathrm{CF}_q}) + \mu_{q,\ell}(\mathsf{RW},\mathrm{CF}_q\wedge\overline{\mathrm{NTV}_q}) \quad (5)\\ &= \max_{\mathcal{B}\in\mathbb{T}} P^{\mathcal{B}\diamond\mathsf{RW}}(\exists\text{distinct } i,j\in\{1,\ldots,q\} : U_i=U_j, S_i=S_j, M_i\neq M_j)\\ &+ \max_{\mathcal{B}\in\mathbb{T}} P^{\mathcal{B}\diamond\mathsf{RW}}(\exists\text{distinct } i,j,k\in\{1,\ldots,q\} : U_i=U_j, S_j=S_k, M_i\neq M_j\neq M_k),\\ &\le \sum_{1\le i<j\le q} \max_{\mathcal{B}\in\mathbb{T}} P^{\mathcal{B}\diamond\mathsf{RW}}(U_i=U_j, H(M_i)=H(M_j), M_i\neq M_j)\\ &+ \sum_{\substack{\text{distinct } i,j,k\\ \in\{1,\ldots,q\}}} \max_{\mathcal{B}\in\mathbb{T}} P^{\mathcal{B}\diamond\mathsf{RW}}(U_i=U_j, H(M_j)+U_j=H(M_k)+U_k, M_i\neq M_j\neq M_k), \quad (6)\end{aligned}$$

where the first inequality follows from union bound, the second follows from the definition of CF_q and Proposition 2. Clearly, for any $\mathcal{B} \in \mathbb{T}$ we have

$$\begin{aligned}
&P^{\mathcal{B}\diamond\mathsf{RW}}(U_i = U_j, H(M_i) = H(M_j), M_i \neq M_j),\\
&= P^{\mathcal{B}\diamond\mathsf{RW}}(H(M_i) = H(M_j), U_i = U_j | M_i \neq M_j) \cdot P^{\mathcal{B}\diamond\mathsf{RW}}(M_i \neq M_j),\\
&\leq \max_{m_i \neq m_j, |m_i|, |m_j| \leq n\ell} \Pr(H(m_i) \neq H(m_j)) \cdot \frac{1}{2^n} \leq \epsilon(\ell) \cdot \frac{1}{2^n}, \qquad (7)
\end{aligned}$$

as $\mathcal{B}$ is T-independent (thus U_i, U_j, M_i, M_j are independent of H's key) and H is $\epsilon(\ell)$-AXU, and that U_i, U_j are uniformly random. In addition, we observe that

$$\begin{aligned}
&P^{\mathcal{B}\diamond\mathsf{RW}}(U_i = U_j, H(M_j) + U_j = H(M_k) + U_k, M_i \neq M_j \neq M_k)\\
&= P^{\mathcal{B}\diamond\mathsf{RW}}(H(M_j) + U_j = H(M_k) + U_k | U_i = U_j, M_i \neq M_j \neq M_k)\\
&\cdot P^{\mathcal{B}\diamond\mathsf{RW}}(M_i \neq M_j \neq M_k | U_i = U_j) \cdot P^{\mathcal{B}\diamond\mathsf{RW}}(U_i = U_j), \qquad (8)\\
&\leq \max_{\substack{m_i \neq m_j \neq m_k, u_j, u_k,\\ |m_i|, |m_j|, |m_k| \leq n\ell}} \Pr(H(m_j) + u_j = H(m_k) + u_k) \cdot \frac{1}{2^n} \leq \epsilon(\ell) \cdot \frac{1}{2^n} \qquad (9)
\end{aligned}$$

from the same reason as above[8]. From Eqs. (4) to (9), we have

$$\nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{NTV}_q}) \leq \left(\binom{q}{3} + \binom{q}{2}\right) \frac{\epsilon(\ell)}{2^n} = (q^3 - q)\frac{\epsilon(\ell)}{6 \cdot 2^n}. \qquad (10)$$

Analysis of the Second Term. Clearly, the probability of $\overline{\mathrm{EQS}(\alpha)}$ is bounded as

$$\begin{aligned}
\nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{EQS}(\alpha)}) &\leq \Pr(^{\exists}\text{distinct } i_1, i_2, \ldots, i_{\alpha+1} : U_{i_1} = U_{i_2} = \cdots = U_{i_{\alpha+1}})\\
&\leq \binom{q}{\alpha+1}\frac{1}{2^{n\alpha}}. \qquad (11)
\end{aligned}$$

Analysis of the First Term. We have the following lemma.

Lemma 1. *If* $\nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q}) \leq 1/2$,

$$\nu_{q,q_v,\ell}(\mathsf{EH}, \mathrm{SUC} | \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q) \leq q_v \left(2\alpha\epsilon(\ell) + \frac{1}{2^\pi}\right).$$

The proof of Lemma 1 is in Appendix D.

Combining Terms. From Eqs. (2), (10), (11), and Lemma 1, $\mathrm{FP}_{\mathsf{EH}}(q, q_v, \ell) \leq \binom{q}{\alpha+1}\frac{1}{2^{n\alpha}} + (q^3 - q)\frac{\epsilon(\ell)}{6 \cdot 2^n} + q_v\left(2\alpha\epsilon(\ell) + \frac{1}{2^\pi}\right)$ for any positive integer $\alpha \geq 2$, if $\nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q}) \leq \binom{q}{\alpha+1}\frac{1}{2^{n\alpha}} + (q^3 - q)\frac{\epsilon(\ell)}{6 \cdot 2^n} \leq 1/2$. By setting $\alpha = 2$ we conclude the proof.

[8] At a glance, $p = P^{\mathcal{B}\diamond\mathsf{RW}}(H(M_j) + U_j = H(M_k) + U_k | U_i = U_j, M_i \neq M_j \neq M_k)$ seems $1/2^n$ irrespective of H as U_i, U_j, and U_k are independent and uniform. This is wrong if $i < k < j$ and H is (e.g.) identity function for n-bit inputs: by choosing $M_k = U_i$ and $M_j = U_k$, p is 1. Moreover p is 1 if H is (a special class of) AU but not AXU. Thus being AU is not the sufficient condition for H.

6 Blockcipher-based Instantiations

6.1 A CBC-based Mode

The generality of our EHtM allows us to derive various concrete instantiations. Here, we present two blockcipher modes of operation. They look similar to RMAC [16]. However they are provably secure on the pseudorandomness of the blockcipher whereas RMAC needs the ideal-cipher model (ICM). Our modes use CBC-MAC and a collision-free message padding, $\mathtt{pad} : \Sigma^* \to \bigcup_{i=0,1,\ldots}(\Sigma^n)^i$. For input x, $\mathtt{pad}$ appends $10^{|x| \bmod n-1}$ to x if $|x| \bmod n \neq 0$, otherwise appends 10^{n-1}, then partitions the appended x into n-bit blocks. For empty string ϕ, we define $\mathtt{pad}(\phi) = 10^{n-1}$. Let $\mathrm{CBC}[E_K] : \bigcup_{i=1,\ldots}(\Sigma^n)^i \to \Sigma^n$ be CBC-MAC using $E_K : \Sigma^n \to \Sigma^n$. For $x = (x_1, \ldots, x_\ell) \in (\Sigma^n)^\ell$, $\mathrm{CBC}[E_K](x) = Y_\ell$, where $Y_i = E_K(x_i \oplus Y_{i-1})$ for $i \geq 1$ and $Y_0 = 0^n$.

Our first proposal, MAC-R1, uses two blockcipher keys and is as follows.

Definition 6. *The mode* MAC-R1 *generates the π-bit tag, T, for message $M \in \Sigma^*$, using $(n-1)$-bit random IV, U, as $T = \mathtt{chop}_\pi(E_{K_2}(U\|0) \oplus E_{K_2}(S\|1))$, where S denotes $U \oplus \mathtt{chop}_{n-1}(\mathrm{CBC}[E_{K_1}](\mathtt{pad}(M)))$. Here K_1 and K_2 are two keys of an n-bit blockcipher, E_K.*

Fig. 2 depicts MAC-R1, where an internal chop is substituted with a logical OR. One may wonder if this really keeps the security beyond the birthday bound, as the use of PRP-PRF switching lemma will bring $O(q^2/2^n)$ into the bound. However, this problem is circumvented by the use of Bernstein's lemma [7] instead of the switching lemma[9]. The security bound of MAC-R1 is as follows.

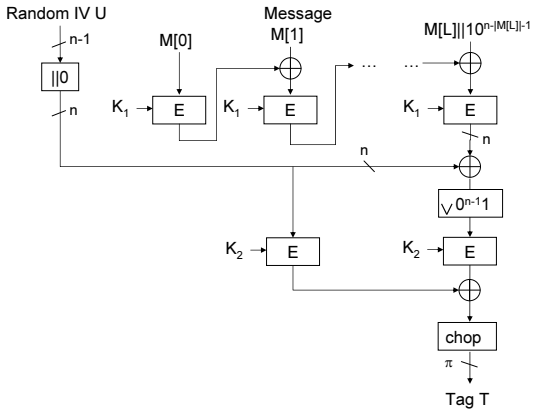

Fig. 2. MAC-R1 when the last message block is partial.

[9] Bernstein's lemma is useful to derive a bound for the ratio (rather than the difference) of two game probabilities where one involves URP and the other involves URF.

Corollary 1. *Let* $\epsilon_{\mathrm{cbc}}(\ell) \stackrel{\mathrm{def}}{=} 2\mathtt{d}(\ell+1)/2^n + 64(\ell+1)^4/2^{2n}$*, where* $\mathtt{d}(x)$ *denotes the maximum number of positive integers that divide* h*, for all* $h \le x$*. Let* $\delta(a) \stackrel{\mathrm{def}}{=} \left(1 - \frac{a-1}{2^n}\right)^{-\frac{a}{2}}$*. Then, we have*

$$\mathrm{FP}_{\text{MAC-R1}[E_{K_1},E_{K_2}]}(q,q_v,\ell,\tau) \le 2\mathtt{Adv}^{\mathtt{prp}}_E(q_1^*,\tau+O(q+q_v)) + \left\{\frac{q^3}{3}\left(\frac{2\epsilon_{\mathrm{cbc}}(\ell+1)}{2^n}+\frac{4}{2^{3n}}\right)+q_v\left(8\epsilon_{\mathrm{cbc}}(\ell+1)+\frac{1}{2^\pi}\right)\right\}\cdot\delta(q_2^*),$$

where $q_1^* = (q+q_v)(\ell+1)$*,* $q_2^* = 2(q+q_v)$*, if* $q^3 \le 1.5\left(\frac{2\epsilon_{\mathrm{cbc}}(\ell+1)}{2^n}+\frac{4}{2^{3n}}\right)^{-1}$.

Proof. Let $\mathsf{P}_n^{(1)}$ and $\mathsf{P}_n^{(2)}$ be independent n-bit URPs. Using Bernstein's lemma (Theorem 2.2 of [7]), we have

$$\mathrm{FP}_{\text{MAC-R1}[\mathsf{P}_n^{(1)},\mathsf{P}_n^{(2)}]}(q,q_v,\ell) \le \mathrm{FP}_{\mathsf{R1PR}}(q,q_v,\ell)\cdot\delta(q_2^*), \tag{12}$$

where R1PR denotes MAC-R1$[\mathsf{P}_n^{(1)},\mathsf{R}_{n,n}]$ (recall $\mathsf{R}_{n,n}$ is an n-bit block URF). As a pair of functions $(\mathsf{R}_{n,n}(*\|0),\mathsf{R}_{n,n}(*\|1))$ is equivalent to a pair of independent URFs : $\Sigma^{n-1} \to \Sigma^n$, R1PR is a complete instantiation of EHtM with $(n-1)$-bit random IV (and hash value). We then need to analyze the hash function of R1PR, namely $H_{\mathsf{R1PR}} = \mathtt{chop}_{n-1} \circ \mathrm{CBC}[\mathsf{P}_n^{(1)}] \circ \mathtt{pad}$. From Bellare et al. [4] and its extension [24], CBC$[\mathsf{P}_n]$ is $\epsilon_{\mathrm{cbc}}(\ell)$-AXU, and thus H_{R1PR} is $2\epsilon_{\mathrm{cbc}}(\ell+1)$-AXU. Combining this observation and Theorem 2 proves that $\mathrm{FP}_{\mathsf{R1PR}}(q,q_v,\ell)$ is at most $\frac{q^3}{3}\left(\frac{2\epsilon_{\mathrm{cbc}}(\ell+1)}{2^n}+\frac{4}{2^{3n}}\right)+q_v\left(8\epsilon_{\mathrm{cbc}}(\ell+1)+\frac{1}{2^\pi}\right)$. From this and Eq. (12), we prove the information-theoretic version of Corollary 1. The computational counterpart is easy.

Inside the Bound. We confirmed that $\delta(q_2^*)$ is well approximated via the first-order approximation, $(1+(q_2^*)^2/2^{n+1})$, when $q_2^* \le 2^{n/2}$. Thus MAC-R1's bound is about $q^3\epsilon_{\mathrm{cbc}}(\ell)/2^n + q_v(\epsilon_{\mathrm{cbc}}(\ell)+1/2^\pi)$ when $q+q_v \le 2^{n/2-1}$. Here, $\epsilon_{\mathrm{cbc}}(\ell)$ grows much slower than $\ell/2^n$ (see [4]). When q_2^* exceeds $2^{n/2}$, $\delta(q_2^*)$ rapidly grows and the bound quickly reaches 1. From this, the bound is almost 1 when $q = 2^{n/2+c}$ for a small positive constant c. This seemingly contradicts with our proposition, but the bound is still negligibly small when $q = 2^{n/2}$. This can be verified by numerical results given in Fig. 3.

6.2 CBC-based, More Secure Mode

As mentioned, the bound of MAC-R1 quickly reaches one as q exceeds $2^{n/2}$. To overcome this problem, we consider a different finalization : $(\Sigma^{n-2})^2 \to \Sigma^n$ as

$$\mathrm{DTWIN}[E_K](x,x') \stackrel{\mathrm{def}}{=} E_K(x\|00) \oplus E_K(x\|10) \oplus E_K(x'\|01) \oplus E_K(x'\|11). \tag{13}$$

Definition 7. *The mode* MAC-R2 *generates the* π*-bit tag,* T*, for message* $M \in \Sigma^*$*, using* $(n-2)$*-bit random IV,* U*, as* $T = \mathtt{chop}_\pi(\mathrm{DTWIN}[E_{K_2}](U,S))$*, where* S *is* $n-2$ *bits and defined as* $U \oplus \mathtt{chop}_{n-2}(\mathrm{CBC}[E_{K_1}](\mathtt{pad}(M)))$.

To derive a bound, we define $\mathrm{TWIN}[E_K] : \Sigma^{n-1} \to \Sigma^n$ as $\mathrm{TWIN}[E_K](x) = E_K(x\|0) \oplus E_K(x\|1)$. Here $\mathrm{DTWIN}[E_K](U, S)$ corresponds to $\mathrm{TWIN}[E_K](U\|0) \oplus \mathrm{TWIN}[E_K](S\|1)$, and Lucks [20] proved $\mathtt{Adv}^{\mathtt{prf}}_{\mathrm{TWIN}[\mathsf{P}_n]}(q) \le 4q/2^n + q^3/3 \cdot 2^{2n-1}$. Hence, the concrete bound of MAC-R2 can be derived without Bernstein's lemma, which is as follows.

Corollary 2.

$$\mathrm{FP}_{\text{MAC-R2}[E_{K_1}, E_{K_2}]}(q, q_v, \ell, \tau) \le 2\mathtt{Adv}^{\mathtt{prp}}_{E}(2q_1^*, \tau + O(q + q_v))$$
$$+ \frac{q^3}{3}\left(\frac{8\epsilon_{\mathrm{cbc}}(\ell+1)}{2^n} + \frac{64}{2^{3n}}\right) + q_v\left(16\epsilon_{\mathrm{cbc}}(\ell+1) + \frac{1}{2^\pi}\right) + \frac{8(q+q_v)}{2^n} + \frac{16(q+q_v)^3}{3 \cdot 2^{2n}}$$

if $q^3 \le 1.5\left(\frac{2\epsilon_{\mathrm{cbc}}(\ell)}{2^n} + \frac{4}{2^{3n}}\right)^{-1}$, *where* $\epsilon_{\mathrm{cbc}}(\ell)$ *and* q_1^* *are as defined by Corollary 1.*

From Corollary 2, the dominant term of MAC-R2's bound is $\epsilon_{\mathrm{cbc}}(\ell)q^3/2^n$ (without the restriction $q + q_v < 2^{n/2-1}$). Thus, MAC-R2 provides the same level of security as that of EHtM with n-bit PRFs.

6.3 A Detailed Comparison

Table. Table 2 presents a detailed comparison of MAC-R1, MAC-R2, and previous MAC modes. Presenting the table is not a straightforward task because of the differences in MAC types, security notions, and parameters. We tried to do a fair comparison while keeping the simplicity. We chose CMAC (a.k.a. OMAC [13]), RMAC, EMAC [10], and MAC-R1 and MAC-R2 with $\pi = n$, where n-bit blockcipher is used. The bounds are shown without minor terms. For CMAC and EMAC, only their prf-advantages are published [4][13][14]. For them we have used Proposition 7.3 of [2] to get the bounds of FP. RMAC has several versions, and we employ one defined in [16]. The RMAC proof is based on the ideal-cipher model. For CMAC and RMAC, the bounds using σ (total message blocks of queries) are also known. As $\sigma \le \ell(q+q_v)$ holds we can always translate a bound using σ into one using (ℓ, q, q_v). The difference is small unless message length distribution has very long tails. We note that one call (two calls) of blockcipher in MAC-R1 (MAC-R2) can be done only with random IV. Hence, when such precomputation is feasible they will be even faster in practice.

Graph. It is still difficult to see the bound shapes from Table 2. Hence, we also perform exact bound computations for $n = 64$ and 128. The $\log_2 \mathrm{FP} - \log_2 q$ graphs are shown in Fig. 3. We assume $q_v = q^{1/2}$, but the bound shape is almost unchanged if q_v is larger, e.g., $q_v = q$. The difference of CMAC and EMAC's bounds is due to the recent advance in the collision analysis of CBC-MAC [4], and will be smaller if ℓ is smaller (or, one can use a result of Nandi [25]). To compute $\mathsf{d}(\ell)$, we used that $\mathsf{d}(\ell) < \lg^2 \ell$ for $\ell < 2^{25}$, shown by [4].

This graph enables us to see how much queries or data are acceptable to restrict the forgery probability being smaller than $2^{-\gamma}$, where γ works as a

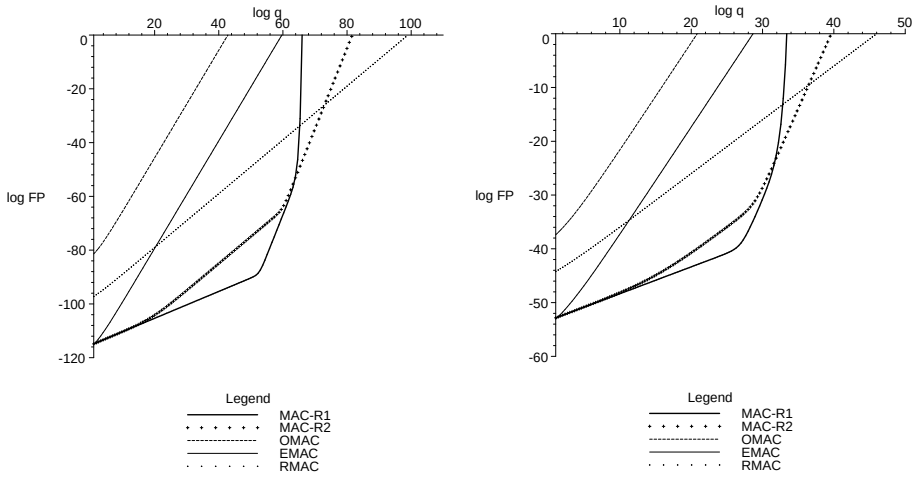

Fig. 3. $\log_2 \mathrm{FP}$ – $\log_2 q$ graphs with $q_v = q^{1/2}$. (left) $n = \pi = 128$, $\ell = 2^{20}$ (right) $n = \pi = 64$, $\ell = 2^{10}$.

security parameter[10]. For example, if we set $\gamma = 20$, the maximum acceptable data amount for $n = 64$ and $\ell = 2^{10}$ is about 14.6 Mbyte for CMAC, 3.2 Gbyte for EMAC, 512.9 Gbyte for RMAC, 40.4 Tbyte for MAC-R1 and 65.6 Tbyte for MAC-R2. In this case, our proposal is even superior to RMAC; it is due to a relatively large constant of RMAC bound ($(4n + 6)\sigma/2^n$ is presented in [16]), and the difference in growths of $q/2^n$ and $q^3/2^{2n}$.

Table 2. Detailed Comparison of MAC Modes.

MAC	Key	Rand	Blockcipher Calls	Security Bound
CMAC	1	–	$\lceil \lvert M\rvert/n\rceil + 1$ (precomp)	$\sigma^2/2^n$ [14] or $\ell^2(q+q_v)^2/2^n$ [13]
EMAC	2	–	$\lceil (\lvert M\rvert + 1)/n\rceil + 1$	$\mathsf{d}(\ell)(q+q_v)^2/2^n$ [4]
RMAC	2	n	$\lceil (\lvert M\rvert + 1)/n\rceil + 1$	$\sigma/2^n$ [16] or $\ell(q+q_v)/2^n$ (with ICM)
MAC-R1	2	$n-1$	$\lceil (\lvert M\rvert + 1)/n\rceil + 2$	$(\mathsf{d}(\ell)q^3/2^{2n} + \mathsf{d}(\ell)q_v/2^n)\cdot\delta(2q+2q_v)$
MAC-R2	2	$n-2$	$\lceil (\lvert M\rvert + 1)/n\rceil + 4$	$(\mathsf{d}(\ell)q^3 + q_v^3)/2^{2n} + (q + \mathsf{d}(\ell)q_v)/2^n$

[10] If we say "it has b-bit security" or "it is secure if $q \ll 2^b$", we implicitly assume $\gamma = 0$. This is a simple, conventional way. However, it is sometimes too weak to grasp the actual values: $q^2/2^n$ can be much smaller than $q/2^{n/2}$ but both mean $n/2$-bit security.

Acknowledgments

We would like to thank Liang Bo and anonymous referees for helpful comments that improved the paper.

References

1. M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. "A Concrete Security Treatment of Symmetric Encryption." *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, FOCS '97, pp. 394-403, 1997.
2. M. Bellare, O. Goldreich, and A. Mityagin. "The Power of Verification Queries in Message Authentication and Authenticated Encryption." *Cryptology ePrint Archive*, 2004/309.
3. M. Bellare, O. Goldreich, and K. Krawczyk. "Stateless Evaluation of Pseudorandom Functions: Security Beyond the Birthday Barrier." *Advances in Cryptology*- CRYPTO '99, LNCS 1666, pp. 270-287, 1999.
4. M. Bellare, K. Pietrzak, and P. Rogaway. "Improved Security Analyses for CBC MACs." *Advances in Cryptology* - CRYPTO '05, LNCS 3621, pp. 527-541, 2005.
5. D. J. Bernstein. "The Poly1305-AES Message-Authentication Code." *Fast Software Encryption*, FSE'05, LNCS 3557, pp. 32-49, 2005.
6. D. J. Bernstein. "Stronger Security Bounds for Wegman-Carter-Shoup Authenticators." *Advances in Cryptology*- EUROCRYPT '05, LNCS 3494, pp. 164-180, 2005.
7. D. J. Bernstein. "Stronger Security Bounds for Permutations." Available from `http://cr.yp.to/papers.html`.
8. J. Black and M. Cochran. "MAC Reforgeability." *Fast Software Encryption*, FSE'09, LNCS 5665, pp.345-362, 2009.
9. J. Black. "Message Authentication Code." *PhD dissertation*, 2000.
10. A. Bosselaers and B. Preneel, *Integrity Primitives for Secure Information Systems, Final Report of RACE Integrity Primitives Evaluation RIPE-RACE*, LNCS 1007, 1995.
11. L. Carter and M. Wegman. "Universal Classes of Hash Functions." *Journal of Computer and System Science*, Vol. 18, pp. 143-154, 1979.
12. Y. Dodis and K. Pietrzak. "Improving the Security of MACs Via Randomized Message Preprocessing." *Fast Software Encryption*-, FSE'07, LNCS 4593, pp.414-433, 2007.
13. T. Iwata and K. Kurosawa. "OMAC: One-Key CBC MAC." *Fast Software Encryption*- FSE'03, LNCS 2887, pp. 129-153, 2003.
14. T. Iwata and K. Kurosawa. "Stronger Security Bounds for OMAC, TMAC, and XCBC." *Progress in Cryptology*- INDOCRYPT'03, LNCS 2904, pp. 402-415, 2003.
15. T. Iwata. "New Blockcipher Modes of Operation with Beyond the Birthday Bound Security." *Fast Software Encryption*- FSE'06, LNCS 4047, pp. 310-327, 2006.
16. E. Jaulmes, A. Joux, and F. Valette. "On the Security of Randomized CBC-MAC Beyond the Birthday Paradox Limit: A New Construction." *Fast Software Encryption*- FSE'02, LNCS 2365, pp. 237-251, 2002.
17. E. Jaulmes and R. Lercier. "FRMAC, a Fast Randomized Message Authentication Code." *Cryptology ePrint Archive*- 2004/166.
18. L. R. Knudsen and T. Kohno. "Analysis of RMAC." *Fast Software Encryption*- FSE'03, LNCS 2887, pp. 182-191, 2003.

19. T. Krovetz "Message Authentication on 64-Bit Architectures." *Selected Areas in Cryptography*- SAC'06, LNCS 4356, pp. 327-341, 2007.
20. S. Lucks. "The Sum of PRPs Is a Secure PRF." *Advances in Cryptology*- EUROCRYPT'00, LNCS 1807, pp. 470-484, 2000.
21. U. Maurer. "Indistinguishability of Random Systems." *Advances in Cryptology*- EUROCRYPT'02, LNCS 2332, pp. 110-132, 2002.
22. D. McGrew and J. Viega. "The Security and Performance of the Galois/Counter Mode (GCM) of Operation." *Progress in Cryptology*- Indocrypt'04, LNCS 3348, pp. 343-355, 2004.
23. D. McGrew and S. Fluhrer. "Multiple forgery attacks against Message Authentication Codes." *Cryptology ePrint Archive*, 2005/161.
24. K. Minematsu and T. Matsushima. "New Bounds for PMAC, TMAC, and XCBC." *Fast Software Encryption*- FSE'07, LNCS 4593, pp.434-451, 2007.
25. M. Nandi. "Improved security analysis for OMAC as a pseudorandom function." Journal of Mathematical Cryptology. Volume 3, Issue 2, pp. 133-148, 2009.
26. M. Semanko. "L-collision Attacks against Randomized MACs." *Advances in Cryptology*- CRYPT'00, LNCS 1880, pp. 216-228, 2000.
27. K. Yasuda. "A One-Pass Mode of Operation for Deterministic Message Authentication- Security beyond the Birthday Barrier." *Fast Software Encryption*, FSE'08, LNCS 5086, pp. 316-333, 2008.
28. M. Wegman and L. Carter. "New Hash Functions and Their Use in Authentication and Set Equality." *Journal of Computer and System Sciences*, Vol. 22, pp. 265-279, 1981.
29. Comments on Draft RMAC Specification, `http://csrc.nist.gov/groups/ST/toolkit/BCM/comments.html`

A Proof of Theorem 1

We abbreviate RWMAC$[H, \mathsf{R}_{2n,n}]$ to RW'. Let $U \in \Sigma^n$ be the random value, and let $V = H(M) \in \Sigma^n$ be the hash value for message M. Then it is trivial to see that the uniqueness of (U_i, V_i) for all $i \in \mathcal{Q}$ (see Sect. 5 for definition of $\mathcal{Q}$), denoted by CF'_q, provides the uniform distribution of tags, $T^q \in (\Sigma^\pi)^q$. From this, we easily obtain

$$\begin{aligned}
&\mathrm{FP}_{\mathsf{RW}'}(q, q_v, \ell) \\
&\le \nu_{q,q_v,\ell}(\mathsf{RW}', \mathrm{SUC} | \mathrm{CF}'_q \wedge \mathrm{EQS}(\alpha)) + \nu_{q,\ell}(\mathsf{RW}', \overline{\mathrm{CF}'_q}) + \nu_{q,\ell}(\mathsf{RW}', \overline{\mathrm{EQS}(\alpha)}) \\
&\le q_v \cdot \nu_{q,1,\ell}(\mathsf{RW}', \mathrm{SUC}_1 | \mathrm{CF}'_q \wedge \mathrm{EQS}(\alpha)) + \mu_{q,\ell}(\mathsf{RW}', \overline{\mathrm{CF}'_q}) + \mu_{q,\ell}(\mathsf{RW}', \overline{\mathrm{EQS}(\alpha)}) \\
&\le q_v \cdot \nu_{q,1,\ell}(\mathsf{RW}', \mathrm{SUC}_1 | \mathrm{CF}'_q \wedge \mathrm{EQS}(\alpha)) + \binom{q}{2} \frac{\epsilon(\ell)}{2^n} + \binom{q}{\alpha+1} \frac{1}{2^{n\alpha}}, \qquad (14)
\end{aligned}$$

where event definitions (SUC, SUC_1, and $\mathrm{EQS}(\alpha)$) and probability definitions (ν and μ) are the same as Sect. 5. For forgery attempt $(\widetilde{U}, \widetilde{M}, \widetilde{T})$, let $\widetilde{V} = H(\widetilde{M})$. We define COL$'$ as the event that $(\widetilde{U}, \widetilde{V}) = (U_i, V_i)$ for some $i \in \mathcal{Q}$. Now we observe

$$\begin{aligned}
\nu_{q,1,\ell}(\mathsf{RW}', \mathrm{SUC}_1 | \mathrm{CF}'_q \wedge \mathrm{EQS}(\alpha)) \le{}& \nu_{q,1,\ell}(\mathsf{RW}', \mathrm{COL}' | \mathrm{CF}'_q \wedge \mathrm{EQS}(\alpha)) \\
&+ \nu_{q,1,\ell}(\mathsf{RW}', \mathrm{SUC}_1 | \overline{\mathrm{COL}'} \wedge \mathrm{CF}'_q \wedge \mathrm{EQS}(\alpha)). \qquad (15)
\end{aligned}$$

Here the last term is $1/2^\pi$ since the real tag for $(\widetilde{U}, \widetilde{M})$ is completely unpredictable given $\overline{\text{COL}'} \wedge \text{CF}'_q \wedge \text{EQS}(\alpha)$ (the same as Eq. (24)). The remaining task is to evaluate the first term of the r.h.s. of Eq. (15). We have

$$\nu_{q,1,\ell}(\mathsf{RW}', \text{COL}'|\text{CF}'_q \wedge \text{EQS}(\alpha)) = \mu_{q,1,\ell}(\mathsf{RW}', \text{COL}'|\text{CF}'_q \wedge \text{EQS}(\alpha)), \tag{16}$$

$$\leq \frac{\mu_{q,1,\ell}(\mathsf{RW}', \text{COL}'|\text{EQS}(\alpha))}{1 - \mu_{q,1,\ell}(\mathsf{RW}', \overline{\text{CF}'_q \wedge \text{EQS}(\alpha)})}. \tag{17}$$

We assume the denominator being at least $1/2$. The numerator is clearly at most $\alpha \cdot \epsilon(\ell)$ as the target forgers are T-independent and any U_i's equivalent class is of size at most α. Thus, we have

$$\text{FP}_{\mathsf{RW}'}(q, q_v, \ell) \leq \binom{q}{2}\frac{\epsilon(\ell)}{2^n} + \binom{q}{\alpha+1}\frac{1}{2^{n\alpha}} + q_v\left(2\alpha \cdot \epsilon(\ell) + \frac{1}{2^\pi}\right), \tag{18}$$

if $\binom{q}{2}\frac{\epsilon(\ell)}{2^n} + \binom{q}{\alpha+1}\frac{1}{2^{n\alpha}} \leq 1/2$, for any $\alpha \geq 2$. If $q \leq 2^{n-2}$ and $\alpha = n-1$, we have $\binom{q}{\alpha+1}\frac{1}{2^{\alpha}} < \frac{1}{2^n}$. Thus, the above implies $q^2\epsilon(\ell)/2^{n+1} + q_v\left(2(n-1)\cdot\epsilon(\ell) + 1/2^\pi\right)$ when $q \leq \min\{2^{n-2}, \sqrt{2^n \cdot \epsilon(\ell)^{-1}}\}$. This concludes the information-theoretic part of the proof. The computational part is trivial.

B Proof of Proposition 1

For simplicity, we assume $\pi = n$ and $\mathcal{Q} = \{1, \ldots, q\}$ (i.e., all (u_i, m_i)s are distinct) throughout the proof; proving under this setting is enough to prove other settings. Let $\mathsf{FNL} : (\Sigma^n)^2 \to \Sigma^n$ be the finalization of EH, i.e. $\mathsf{FNL}(u, s) = \mathsf{R}^{(1)}(u) \oplus \mathsf{R}^{(2)}(s)$. Note that $\mathsf{FNL}(u, s)$ is equivalent to $\mathsf{R}_{n+1,n}(U\|0) \oplus \mathsf{R}_{n+1,n}(S\|1)$. Then, a pair of CV $(\lambda(U), \lambda(S))$ can be expressed as $\Lambda(U, S) = \lambda(U\|0) \oplus \lambda(S\|1)$, where $\lambda(U\|0)$ and $\lambda(S\|1)$ are 2^{n+1} bits CVs. Here $\Lambda(U, S)$'s weight is always 2, as $U\|0$ and $S\|1$ never collide. Then, from Sect. 5.2 of [21] (or [3]), when the set $\{\Lambda(U_1, S_1), \ldots, \Lambda(U_q, S_q)\}$ is linearly independent the outputs of FNL are perfectly random. Since this condition is equivalent to LID_q, we have

$$P^{\mathsf{FNL}}(T_q = t_q | U^q = u^q, S^q = s^q, T^{q-1} = t^{q-1}, \mathcal{E}) = 1/2^n \tag{19}$$

for all possible arguments, when $\mathcal{E} = \text{LID}_q$. When $\mathcal{E} = \text{NTV}_q$, Eq. (19) also holds since $\text{NTV}_q \to \text{LID}_q$ and that NTV_q is defined over (U^q, S^q) as well as LID_q. As T_q's distribution in Eq. (19) is independent of actual values of U^q and S^q, we immediately obtain

$$P^{\mathsf{EH}}(T_q = t_q | U^q = u^q, M^q = m^q, T^{q-1} = t^{q-1}, \mathcal{E}) = 1/2^n \tag{20}$$

for all possible arguments, and for both $\mathcal{E} = \text{LID}_q$ and NTV_q. This proves Eq. (3) for $\mathsf{Func} = \mathsf{EH}$. When $\mathsf{Func} = \mathsf{RW}$, the proof follows from the fact that both NTV_q and LID_q includes CF_q, which assures the uniform distribution of T_q given $U^q = u^q, M^q = m^q$, and $T^{q-1} = t^{q-1}$.

We proceed to the proof of Eq. (4). From Eq. (19) it is clear that

$$P^{\mathsf{FNL}}(T_q = t_q, \mathrm{NTV}_q | U^q = u^q, S^q = s^q, T^{q-1} = t^{q-1}, \mathrm{NTV}_{q-1}) \\ = P^{\mathsf{R}_{2n,n}}(T_q = t_q, \mathrm{NTV}_q | U^q = u^q, S^q = s^q, T^{q-1} = t^{q-1}, \mathrm{NTV}_{q-1}) \quad (21)$$

holds for all possible arguments (recall that we assumed unique (u_q, s_q)). From Lemma 4 of [21], this equality also holds true for $\mathsf{FNL} \circ \mathsf{Pre}$ and $\mathsf{R}_{2n,n} \circ \mathsf{Pre}$, for any independently-keyed pre-processing $\mathsf{Pre} : (\Sigma^n)^2 \to (\Sigma^n)^2$. Thus, by defining the pre-processing as $(U, M) \to (U, U \oplus H(M))$, we obtain

$$P^{\mathsf{EH}}(T_q = t_q, \mathrm{NTV}_q | U^q = u^q, M^q = m^q, T^{q-1} = t^{q-1}, \mathrm{NTV}_{q-1}) \\ = P^{\mathsf{RW}}(T_q = t_q, \mathrm{NTV}_q | U^q = u^q, M^q = m^q, T^{q-1} = t^{q-1}, \mathrm{NTV}_{q-1}). \quad (22)$$

From Lemma 6 of [21], the above implies $\nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{NTV}_q}) = \nu_{q,\ell}(\mathsf{RW}, \overline{\mathrm{NTV}_q})$. In RW, the q tags are independently random as long as NTV_q is satisfied, thus the maximum probability of $\overline{\mathrm{NTV}_q}$ can be achieved without seeing tags, that is, by T-independent forgers (this is a simple extension of Corollary 1 (iv) of [21]: the difference is that Corollary 1 (iv) of [21] only considers chosen inputs with no randomness while in our case a part of input is independently random). Therefore, we have $\nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{NTV}_q}) = \nu_{q,\ell}(\mathsf{RW}, \overline{\mathrm{NTV}_q}) = \mu_{q,\ell}(\mathsf{RW}, \overline{\mathrm{NTV}_q})$, which concludes the proof of Eq. (4).

C Proof of Proposition 2

Let $\mathcal{E} = \mathcal{E}_1 \wedge \mathcal{E}_2$, where $\mathcal{E}_1 \stackrel{\mathrm{def}}{=} [\stackrel{\exists}{i, j, k} \in \mathcal{Q}, U_i = U_j \neq U_k \wedge S_i \neq S_j = S_k]$ and $\mathcal{E}_2 \stackrel{\mathrm{def}}{=} [\forall i', j' \in \mathcal{Q}, (U_{i'}, S_{i'}) \neq (U_{j'}, S_{j'})]$. Note that $\mathcal{E}_2 \equiv \mathrm{CF}_q$. Also, it is easy to see that $\mathcal{E}_1$ is equivalent to $[\stackrel{\exists}{i, j, k} \in \{1, \ldots, q\}, U_i = U_j \neq U_k \wedge S_i \neq S_j = S_k, M_i \neq M_j \neq M_k]$. Using this, what we need to prove is $\mathrm{CF}_q \wedge \overline{\mathrm{NTV}_q} \equiv \mathcal{E}$. This equivalence trivially holds when $|\mathcal{Q}| \leq 2$, as both sides are false in this case. When $\bar{q} \stackrel{\mathrm{def}}{=} |\mathcal{Q}| \geq 3$, w.l.o.g. we assume the set $\{(U_i, M_i)\}_{i=1,\ldots,\bar{q}}$ consists of unique elements (i.e., $\mathcal{Q} = \{1, \ldots, \bar{q}\}$).

If a subset $\mathcal{I} \subseteq \{1, \ldots, \bar{q}\}$ whose size is an odd number ≥ 3 satisfies that $\mathrm{Hw}(\bigoplus_{\mathcal{I}} \lambda(U, S)) = (1, 1)$ and any $\mathcal{I}' \subset \mathcal{I}$ whose size is an odd number ≥ 3 has $\mathrm{Hw}(\bigoplus_{\mathcal{I}'} \lambda(U, S)) \neq (1, 1)$, $\mathcal{I}$ is called the minimal index set. When $\mathrm{CF}_q \wedge \overline{\mathrm{NTV}_q}$ holds, there exists at least one minimal index set, which will be denoted by $\mathcal{I}^*$ (it may not be unique). The set $\{U_i\}_{i \in \mathcal{I}^*}$ is uniquely partitioned into equivalent classes, i.e. the sets of identical elements. We say U_i is odd-colliding (even-colliding) if the size of U_i's equivalent class in $\mathcal{I}^*$ is odd (even). We use the same definition for $\{S_i\}_{i \in \mathcal{I}^*}$. If U_i and S_i are both odd-colliding, we say (U_i, S_i) is an odd-odd pair. In $\{(U_i, S_i)\}_{i \in \mathcal{I}^*}$, there is a unique equivalent class of U whose size is odd, and a unique equivalent class of S whose size is odd, too. Here multiple odd-odd pairs do not exist in $\{(U_i, S_i)\}_{i \in \mathcal{I}^*}$, as this implies $\overline{\mathrm{CF}_q}$. Moreover, any odd-odd pair does not exist; if it exists when $|\mathcal{I}^*| = 3$, $\overline{\mathrm{CF}_q}$ occurs by the remaining two pairs, and when $|\mathcal{I}^*| > 3$ (as $|\mathcal{I}^*|$ must be odd, we have $|\mathcal{I}^*| \geq 5$), removing the unique odd-odd pair and an even-even pair will result

in an index set $\mathcal{I}' \subset \mathcal{I}^*$ satisfying $\mathrm{Hw}(\bigoplus_{\mathcal{I}'} \lambda(U,S)) = (1,1)$, thus contradicting to the minimality of $\mathcal{I}^*$. Therefore, we must have at least one odd-even or even-odd pair in $\mathcal{I}^*$. Let us assume that (U_i, S_i) is such an odd-even pair. As S_i is even-colliding, there exists $j \neq i$, $j \in \mathcal{I}^*$ such that $S_i = S_j$. This implies $U_j \neq U_i$ as $U_j = U_i$ means a (U,S)-collision. From $U_j \neq U_i$, we know U_j is even-colliding, and thus there exists $k \in \mathcal{I}^* \setminus \{i,j\}$ such that $U_j = U_k$. Then $S_k \neq S_j$ holds from CF_q. The case that there exists one even-odd pair holds true from the symmetry. This proves the direct part, $\mathrm{CF}_q \wedge \overline{\mathrm{NTV}_q} \rightarrow \mathcal{E}$. The converse clearly holds true, and thus we have $\mathrm{CF}_q \wedge \overline{\mathrm{NTV}_q} \equiv \mathcal{E}$. From the definition of $\mathcal{E}$, the proof is completed.

D Proof of Lemma 1

First, we have

$$\begin{aligned}\nu_{q,q_v,\ell}(\mathsf{EH}, \mathrm{SUC} | \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q) &\leq q_v \cdot \nu_{q,1,\ell}(\mathsf{EH}, \mathrm{SUC}_1 | \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q), \\ &\leq q_v \cdot \nu_{q,1,\ell}(\mathsf{EH}, \overline{\mathrm{LID}_{q+1}} | \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q) \\ &+ q_v \cdot \nu_{q,1,\ell}(\mathsf{EH}, \mathrm{SUC}_1 | \mathrm{LID}_{q+1} \wedge \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q),\end{aligned} \tag{23}$$

where LID_{q+1} denotes the event that $\{(\lambda(U_i), \lambda(S_i))\}_{i \in \mathcal{Q}} \cup \{(\lambda(\widetilde{U}), \lambda(\widetilde{S}))\}$, where $\widetilde{S} = \widetilde{U} \oplus H(\widetilde{M})$, is linearly independent. Obviously, $(\widetilde{U}, \widetilde{M}) \neq (U_i, M_i)$ for any $i \leq q$. Let $T_{\mathrm{real}} = \mathsf{EH}(\widetilde{U}, \widetilde{M})$ be the real tag for $(\widetilde{U}, \widetilde{M})$. If LID_{q+1} occurs, T_{real} is uniform and independent of previous transcripts. Hence, T_{real} is completely unpredictable. This means that

$$\nu_{q,1,\ell}(\mathsf{EH}, \mathrm{SUC}_1 | \mathrm{LID}_{q+1} \wedge \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q) = 1/2^{\pi}. \tag{24}$$

To see $\nu_{q,1,\ell}(\mathsf{EH}, \overline{\mathrm{LID}_{q+1}} | \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q)$, the occurrence of $\overline{\mathrm{LID}_{q+1}}$ indicates $\mathrm{Hw}(\bigoplus_{\mathcal{I}} \lambda(U) \oplus \lambda(\widetilde{U}), \bigoplus_{\mathcal{I}} \lambda(S) \oplus \lambda(\widetilde{S})) = (0,0)$ for an index set $\mathcal{I} \subseteq \mathcal{Q}$. Thus we have $\mathrm{Hw}(\bigoplus_{\mathcal{I}} \lambda(U), \bigoplus_{\mathcal{I}} \lambda(S)) = (1,1)$. As we have NTV_q in the conditional clause, this is impossible if $|\mathcal{I}| \geq 3$, and also impossible if $|\mathcal{I}| = 2$ (as any index set of even size can not produce $(1,1)$). The only possibility is $|\mathcal{I}| = 1$. This corresponds to the event $\mathrm{COL} \stackrel{\mathrm{def}}{=} [\exists i \in \mathcal{Q}, (\widetilde{U}, \widetilde{S}) = (U_i, S_i)]$. Thus we obtain

$$\nu_{q,1,\ell}(\mathsf{EH}, \overline{\mathrm{LID}_{q+1}} | \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q) = \nu_{q,1,\ell}(\mathsf{EH}, \mathrm{COL} | \mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q). \tag{25}$$

Note that COL is a function of H's key, $(\widetilde{U}, \widetilde{M})$, and (U^q, M^q). When NTV_q occurs, any information on H's key, K, cannot be obtained from T^q, as they are independent of K (from Prop. 1). Thus, the maximum of the conditional probability of COL given NTV_q can be achieved by T-independent forgers. Thus,

by defining $\mathbb{T}$ as the set of all T-independent (q, q_v, ℓ)-forgers, we have[11]

$$\begin{aligned}
&\nu_{q,1,\ell}(\mathsf{EH}, \mathrm{COL}|\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q) = \mu_{q,1,\ell}\left(\mathsf{EH}, \mathrm{COL}|\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q\right) \\
&= \max_{\mathcal{B}\in\mathbb{T}} P^{\mathcal{B}\diamond\mathsf{EH}}\left(\mathrm{COL}|\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q\right) \\
&\leq \max_{\mathcal{B}\in\mathbb{T}} \frac{P^{\mathcal{B}\diamond\mathsf{EH}}(\mathrm{COL}|\mathrm{EQS}(\alpha))}{P^{\mathcal{B}\diamond\mathsf{EH}}(\mathrm{NTV}_q|\mathrm{EQS}(\alpha))}, \\
&\leq \max_{\mathcal{B}\in\mathbb{T}} \frac{P^{\mathcal{B}\diamond\mathsf{EH}}(\mathrm{COL}|\mathrm{EQS}(\alpha))}{1 - P^{\mathcal{B}\diamond\mathsf{EH}}(\overline{\mathrm{NTV}_q \wedge \mathrm{EQS}(\alpha)})}, \qquad (26) \\
&\leq \frac{\max_{\mathcal{B}\in\mathbb{T}} P^{\mathcal{B}\diamond\mathsf{EH}}(\mathrm{COL}|\mathrm{EQS}(\alpha))}{1 - \max_{\mathcal{B}'\in\mathbb{T}} P^{\mathcal{B}'\diamond\mathsf{EH}}(\overline{\mathrm{NTV}_q \wedge \mathrm{EQS}(\alpha)})} \leq \frac{\mu_{q,1,\ell}(\mathsf{EH}, \mathrm{COL}|\mathrm{EQS}(\alpha))}{1 - \nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q})}, \qquad (27)
\end{aligned}$$

as T-independent forger is a subclass of normal forger. If $\mu_{q,1,\ell}(\mathsf{EH}, \mathrm{COL}|\mathrm{EQS}(\alpha))$ is achieved by some $\mathcal{B}^* \in \mathbb{T}$, we have

$$\begin{aligned}
&\mu_{q,1,\ell}(\mathsf{EH}, \mathrm{COL}|\mathrm{EQS}(\alpha)) = P^{\mathcal{B}^*\diamond\mathsf{EH}}(\mathrm{COL}|\mathrm{EQS}(\alpha)) \\
&\leq \sum P^{\mathcal{B}^*\diamond\mathsf{EH}}(\mathrm{COL}|U^q = u^q, M^q = m^q, \widetilde{U} = \widetilde{u}, \widetilde{M} = \widetilde{m}, \mathrm{EQS}(\alpha)) \\
&\cdot P^{\mathcal{B}^*\diamond\mathsf{EH}}(U^q = u^q, M^q = m^q, \widetilde{U} = \widetilde{u}, \widetilde{M} = \widetilde{m}|\mathrm{EQS}(\alpha)) \\
&\leq \max P^{\mathsf{EH}}(^{\exists} i : H(\widetilde{m}) = H(m_i), \widetilde{u} = u_i | U^q = u^q, M^q = m^q, \widetilde{U} = \widetilde{u}, \widetilde{M} = \widetilde{m}), \\
&\leq \max \sum_{i\in\mathcal{Q}:\widetilde{u}=u_i} P^H(H(\widetilde{m}) = H(m_i)) \leq \alpha\epsilon(\ell), \qquad (28)
\end{aligned}$$

where the first sum and two maximums are taken for $(u^q, m^q, \widetilde{u}, \widetilde{m})$ such that (u^q, m^q) satisfies $\mathrm{EQS}(\alpha)$ and $(\widetilde{u}, \widetilde{m}) \neq \forall(u_i, m_i)$. The third inequality follows from that $U^q, M^q, \widetilde{U}$ and $\widetilde{M}$ are independent of H's key (as $\mathcal{B}^*$ is T-independent), and the last inequality follows from that $|\{i : \widetilde{u} = u_i\}| \leq \alpha$ as $\mathrm{EQS}(\alpha)$, and that H is $\epsilon(\ell)$-AXU. From Eqs. (23), (24), (27), (28), we have

$$\nu_{q,q_v,\ell}(\mathsf{EH}, \mathrm{SUC}|\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q) \leq q_v\left(2\alpha\epsilon(\ell) + 1/2^\pi\right), \qquad (29)$$

with the assumption $\nu_{q,\ell}(\mathsf{EH}, \overline{\mathrm{EQS}(\alpha) \wedge \mathrm{NTV}_q}) \leq 1/2$. This concludes the proof.

[11] Here we derive an upper bound of the probability of a "bad" event B conditioned by a "good" event G. For a (randomized) HtM we need a similar analysis where B is the hash collision between verification and tagging queries, and G is the uniqueness of random IVs. Note that, while the uniqueness of random IVs in HtM gives no information on the hash values, the good event G = EQS ∧ NTV for EHtM may give some, negligble information on the hash values. This is the reason why $2\alpha\epsilon(\ell)$ is needed rather than $\alpha\epsilon(\ell)$ in Eq. (29).