Tight Security Bounds for Double-block Hash-then-Sum MACs

Seongkwang Kim, Byeonghak Lee, and Jooyoung Lee*

KAIST, Daejeon, Korea
{ksg0923,lbh0307,hicalf}@kaist.ac.kr

Abstract. In this work, we study the security of deterministic MAC constructions with a double-block internal state, captured by the *double-block hash-then-sum* (DbHtS) paradigm. Most DbHtS constructions, including PolyMAC, SUM-ECBC, PMAC-Plus, 3kf9 and LightMAC-Plus, have been proved to be pseudorandom up to $2^{\frac{2n}{3}}$ queries when they are instantiated with an n-bit block cipher, while the best known generic attacks require $2^{\frac{3n}{4}}$ queries.
We close this gap by proving the PRF-security of DbHtS constructions up to $2^{\frac{3n}{4}}$ queries (ignoring the maximum message length). The core of the security proof is to refine Mirror theory that systematically estimates the number of solutions to a system of equations and non-equations, and apply it to prove the security of the finalization function. Then we identify security requirements of the internal hash functions to ensure $3n/4$-bit security of the resulting constructions when combined with the finalization function.
Within this framework, we prove the security of DbHtS whose internal hash function is given as the concatenation of a universal hash function using two independent keys. This class of constructions include PolyMAC and SUM-ECBC. Moreover, we prove the security of PMAC-Plus, 3kf9 and LightMAC-Plus up to $2^{\frac{3n}{4}}$ queries.

Keywords: message authentication codes, beyond-birthday-bound security, pseudorandom functions, Mirror theory

1 Introduction

MACs. A message authentication code (MAC) is typically built from a block cipher, e.g., CBC-MAC [3], PMAC [5], OMAC [10], LightMAC [13] or from a cryptographic hash function, e.g., HMAC [2]. At a high level, many of these constructions follow the well-established *UHF-then-PRF* design paradigm: a message is first mapped onto a short string through a universal hash function (UHF), and then encrypted through a fixed-input-length PRF to obtain a short tag. This

* Jooyoung Lee was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT), No. NRF-2017R1E1A1A03070248.

method is simple, in particular, being deterministic and stateless, yet its security caps at the so-called birthday bound; any collision at the output of the UHF, which translates into a tag collision, is usually enough to break the security of the scheme. However, the birthday bound security might not be enough, in particular, when the MAC construction is instantiated with a lightweight block cipher such as PRESENT [6], LED [9], GIFT [1] operating on small blocks.

Double-block Hash-then-Sum. Many studies tried to tweak the UHF-then-PRF design in order to obtain beyond-birthday secure MACs, while they possess a similar structural design; the internal state of the hash function is doubled, the two n-bit hash values are encrypted by a block cipher using independent keys, and the outputs are xored to generate the final tag. Datta et al. [7] have dubbed this generic design principle the *double-block hash-then-sum* (DbHtS) paradigm. Within this unified framework, they revisited the security proof of existing DbHtS constructions, including PolyMAC (based on polynomial evaluation [8, 4, 17]), SUM-ECBC [18], PMAC-Plus [19], 3kf9 [20] and LightMAC-Plus [14], and confirmed that all the constructions are secure up to $2^{\frac{2n}{3}}$ queries (ignoring the maximum message length) when they are instantiated with an n-bit block cipher. Recently, Leurent et al. [12] proposed generic attacks on all these constructions using $2^{\frac{3n}{4}}$ (short message) queries, leaving a gap between the upper and lower bounds for the provable security of DbHtS constructions.

Our Results. The goal of this paper is to close this gap by proving the exact PRF-security of DbHtS constructions. In order to do this, we take a modular approach; the first step is to refine Mirror theory [15, 16] that systematically estimates the number of solutions to a system of equations and non-equations in order to prove the security of the finalization function up to $2^{\frac{3n}{4}}$ queries. However, we cannot directly apply Mirror theory to the problem in a black box manner since it requires that $\xi_{max}^2 q \leq 2^n$ in its original form, where ξ_{max} and q denote the maximum component size and the number of equations, respectively. So we refine Mirror theory by distinguishing components of size two and larger ones, and make sharp estimation for components of size two, while we use the fact that the number of larger components is probabilistically small.

The next step is to identify security requirements of the internal hash functions to ensure $3n/4$-bit security of the entire constructions, combined with the finalization function. Existing security proofs limit the probability of having a trail of length 3 in the transcript graph when an adversary makes $2^{\frac{2n}{3}}$ queries, while our proof allows an adversary making $2^{\frac{3n}{4}}$ queries. So in this case, we need to limit the probability of having a trail of length 4 in the transcript graph; this is the most challenging part of the proof (e.g., Lemma 4 for the proof of PMAC-Plus) that needs a careful case analysis.

As a result, we prove the security of various DbHtS constructions including PolyMAC, SUM-ECBC, PMAC-Plus, 3kf9 and LightMAC-Plus up to $2^{\frac{3n}{4}}$ queries, ignoring the maximum message length. Table 1 compares our new bounds to the old ones given in [7]. For some constructions, one cannot simply ignore the influence of the maximum message length on the security bounds. As seen in Figure 1, our bound for PMAC-Plus is better than the old one when ℓ is relatively

small (while our bound is worse for a larger ℓ). So our new bound should be seen as complementary to the old one. However, we also remark that our security proof does not use independent randomness of two masking keys Δ_0 and Δ_1; a single masking key is sufficient for our security proof. We would be able to remove the $\ell^2 q/2^n$ term from the security bound by a more complicated proof using the independent randomness of two masking keys.

Table 1: New security bounds for DbHtS MACs. The number of queries and the maximum message length (in blocks) are denoted q and ℓ, respectively. All the constructions (except PolyMAC) are based on an n-bit block cipher. LightMAC-Plus uses an additional parameter s, which is the size of the prefix for each block cipher call; one can assume $\ell = 2^s - 1$.

Construction	# Keys	Rate	Old bound	New bound
PolyMAC	4	–	$\ell^2 q^3/2^{2n}$	$\ell q^{\frac{4}{3}}/2^n$
SUM-ECBC	4	$\frac{1}{2}$	$\ell^2 q/2^n + q^3/2^{2n}$	$\ell^{o(1)} q^{\frac{4}{3}}/2^n + \ell^4 q^{\frac{4}{3}}/2^{2n}$
PMAC-Plus	3	1	$\ell q^3/2^{2n} + \ell^2 q^2/2^{2n}$	$\ell/2^{\frac{n}{2}} + \ell^{\frac{2}{3}} q^{\frac{4}{3}}/2^n + \ell^2 q/2^n$
3kf9	3	1	$\ell^4 q^3/2^{2n}$	$\ell^{\frac{4}{3}} q^{\frac{4}{3}}/2^n + \ell^2 q^2/2^{2n} + \ell^6 q^4/2^{3n}$
LightMAC-Plus	3	$\frac{n-s}{n}$	$q^3/2^{2n}$	$q^{\frac{4}{3}}/2^n$

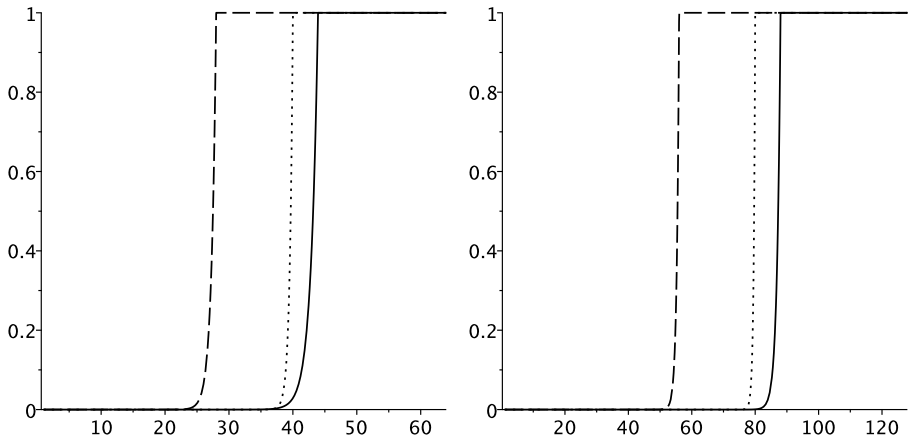
2 Preliminaries

Basic Notation. In all of the following, we fix a positive integer n, and denote $N = 2^n$. We denote 0^n (i.e., n-bit string of all zeros) by $\mathbf{0}$. The set $\{0,1\}^n$ is sometimes regarded as a set of integers $\{0, 1, \ldots, 2^n - 1\}$ by converting an n-bit string $a_{n-1} \cdots a_1 a_0 \in \{0,1\}^n$ to an interger $a_{n-1} 2^{n-1} + \cdots + a_1 2 + a_0$. We also identify $\{0,1\}^n$ with a finite field $\mathbf{GF}(2^n)$ with 2^n elements. For a positive integer q, we write $[q] = \{1, \ldots, q\}$.

Given a non-empty set $\mathcal{X}$, $x \leftarrow_{\$} \mathcal{X}$ denotes that x is chosen uniformly at random from $\mathcal{X}$. 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 all permutations of $\{0,1\}^n$ is simply denoted $\mathsf{Perm}(n)$. The set of all sequences that consist of b pairwise distinct elements of $\mathcal{X}$ is denoted $\mathcal{X}^{*b}$. For integers $1 \le b \le a$, we will write $(a)_b = a(a-1)\cdots(a-b+1)$ and $(a)_0 = 1$ by convention. If $|\mathcal{X}| = a$, then $(a)_b$ becomes the size of $\mathcal{X}^{*b}$.

When two sets $\mathcal{X}$ and $\mathcal{Y}$ are disjoint, their (disjoint) union is denoted $\mathcal{X} \sqcup \mathcal{Y}$. For a set $\mathcal{X} \subset \{0,1\}^n$ and $\lambda \in \{0,1\}^n$, we will write $\mathcal{X} \oplus \lambda = \{x \oplus \lambda : x \in \mathcal{X}\}$.

PRFs and PRPs. Let $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ be a keyed function with key space $\mathcal{K}$, domain $\mathcal{X}$, and range $\mathcal{Y}$, where $\mathcal{X}$ is a subset of $\{0,1\}^*$. We will denote $F_K(X)$ for



(a) $n = 64$ and $\ell = 2^8$ (b) $n = 128$ and $\ell = 2^{16}$

Fig. 1: Upper bounds on distinguishing advantage for PMAC-Plus. The solid and the dotted lines represent the new and the old bounds, respectively. The dashed line represents the security bound $\ell q^2/2^n$ for PMAC. The x-axis gives the log (base 2) of q, and the y-axis gives the security bounds.

$F(K, X)$. A (q, t, ℓ)-distinguisher against F is an algorithm $\mathcal{D}$ with oracle access to a function from $\mathcal{X}$ to $\mathcal{Y}$, making at most q oracle queries, each of length at most ℓ in blocks, running in time at most t, and outputting a single bit. The advantage of $\mathcal{D}$ in breaking the PRF-security of F, i.e., in distinguishing F from a uniformly randomly chosen function $R \leftarrow_{\$} \mathsf{Func}(\mathcal{X}, \mathcal{Y})$, is defined as

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

When $\mathcal{X} = \mathcal{Y}$ and $F(K, \cdot)$ is a permutation for each $K \in \mathcal{K}$, the PRP-security of F is defined as

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

For $\mathsf{atk} \in \{\mathsf{prf}, \mathsf{prp}\}$, we define $\mathbf{Adv}_F^{\mathsf{atk}}(q, t, \ell)$ as the maximum of $\mathbf{Adv}_F^{\mathsf{atk}}(\mathcal{D})$ over all (q, t, ℓ)-distinguishers against F. We will consider PRP-security only for a block cipher whose input size is fixed (e.g., $\mathcal{X} = \{0,1\}^n$); in this case, we will simply drop the parameter ℓ. On the other hand, when we consider information theoretic security, we will drop the parameter t.

Almost Universal Hash Functions. Let $\delta > 0$, and let $H : \mathcal{K}_h \times \mathcal{X} \to \mathcal{Y}$ be a keyed function for three non-empty sets $\mathcal{K}_h$, $\mathcal{X}$, and $\mathcal{Y}$. H is said to be *δ-almost universal* if for any distinct X and $X' \in \mathcal{X}$,

$$\Pr\left[K_h \leftarrow_{\$} \mathcal{K}_h : H_{K_h}(X) = H_{K_h}(X')\right] \leq \delta.$$

Double-block Hash-then-Sum Constructions. Let

$$\begin{aligned} H : \mathcal{K}_h \times \mathcal{M} &\longrightarrow \{0,1\}^n \times \{0,1\}^n \\ (K_h, M) &\longmapsto H_{K_h}(M) \end{aligned}$$

be a keyed function. We will write the $2n$-bit function H as the concatenation of two n-bit functions F and G. So we have

$$H_{K_h}(M) = (F_{K_h}(M), G_{K_h}(M)) .$$

Given a block cipher

$$\begin{aligned} E : \mathcal{K} \times \{0,1\}^n &\longrightarrow \{0,1\}^n \\ (K, X) &\longmapsto E_K(X), \end{aligned}$$

one can define the DbHtS construction based on H and E as follows.

$$\begin{aligned} \mathsf{DbHtS}[H, E] : (\mathcal{K}_h \times \mathcal{K}^2) \times \mathcal{M} &\longrightarrow \{0,1\}^n \\ ((K_h, K_1, K_2), M) &\longmapsto E_{K_1}(F_{K_h}(M)) \oplus E_{K_2}(G_{K_h}(M)). \end{aligned}$$

In a typical MAC based on an n-bit block cipher, the message space is given as the set of all binary strings, namely $\{0,1\}^*$, and a padding scheme

$$\mathsf{pad} : \{0,1\}^* \longrightarrow \bigcup_{i=1}^{\infty} (\{0,1\}^n)^i$$

is used, where pad is a public injective function. Since the padding scheme does not affect the PRF-security of its MAC, we will simply assume that

$$\mathcal{M} = \bigcup_{i=1}^{\ell} (\{0,1\}^n)^i ,$$

where ℓ denotes the maximum message length in blocks (after padding).

H-coefficient Technique. Consider the DbHtS construction based on H and E using keys $K = (K_h, K_1, K_2)$. The first step of the security proof is to replace the keyed permutations E_{K_1} and E_{K_2} by independent random permutations; the resulting construction will be denoted $\mathsf{DbHtS}[H]$ instead of $\mathsf{DbHtS}[H, E]$.

Suppose that a distinguisher $\mathcal{D}$ adaptively makes q queries to the construction oracle, which is either $\mathsf{DbHtS}[H]_{K_h,\pi_1,\pi_2}$ for a random key $K_h \in \mathcal{K}_h$ and independent random permutations π_1 and π_2 (in the real world) or a truly random function R (in the ideal world), recording all the queries $(M_i, T_i)_{1 \le i \le q}$. So according to the instantiation, it would imply either $\mathsf{DbHtS}[H]_{K_h,\pi_1,\pi_2}(M_i) = T_i$ or $R(M_i) = T_i$.

At the end of the interaction, we will give K_h to $\mathcal{D}$ for free. In the ideal world, a dummy key K_h will be selected uniformly at random from $\mathcal{K}_h$, and

given to $\mathcal{D}$. This will not degrade the adversarial distinguishing advantage since the distinguisher is free to ignore this additional information. We will call

$$\tau = (K_h, (M_1, T_1), \ldots, (M_q, T_q))$$

the *transcript* of the attack; it contains all the information that $\mathcal{D}$ has obtained at the end of the attack. We will assume that $\mathcal{D}$ is information theoretic, so we can further assume that $\mathcal{D}$ is deterministic without making any redundant query.

A transcript τ is called *attainable* if the probability to obtain this transcript in the ideal world is non-zero. Any key $K_h \in \mathcal{K}_h$ and any sequence $(T_1, \ldots, T_q) \in (\{0,1\}^n)^q$ uniquely determine an attainable transcript $\tau = (K_h, (M_i, T_i)_{1 \leq i \leq q}))$ containing them, for some $(M_i) \in (\{0,1\}^n)^q$. We denote Γ the set of attainable transcripts. We also denote T_{re} (resp. T_{id}) the probability distribution of the transcript τ induced by the real world (resp. the ideal world). By extension, we use the same notation to denote a random variable distributed according to each distribution.

In order to upper bound the advantage of the distinguisher, we will partition the set of attainable transcripts Γ into a set of "good" transcripts Γ_{good} such that the probabilities to obtain some transcript $\tau \in \Gamma_{\mathsf{good}}$ are close in the real and in the ideal world, and a set Γ_{bad} of "bad" transcripts such that the probability to obtain any $\tau \in \Gamma_{\mathsf{bad}}$ is small in the ideal world, and use the following theorem.

Lemma 1. *Fix a distinguisher $\mathcal{D}$. Let $\Gamma = \Gamma_{\mathsf{good}} \sqcup \Gamma_{\mathsf{bad}}$ be a partition of the set of attainable transcripts. Assume that there exists ε_1 such that for any $\tau \in \Gamma_{\mathsf{good}}$,*

$$\frac{\Pr\left[\mathsf{T}_{\mathrm{re}} = \tau\right]}{\Pr\left[\mathsf{T}_{\mathrm{id}} = \tau\right]} \geq 1 - \varepsilon_1,$$

and that there exists ε_2 such that $\Pr[\mathsf{T}_{\mathrm{id}} \in \Gamma_{\mathsf{bad}}] \leq \varepsilon_2$. Then one has

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{DbHtS}[H]}(\mathcal{D}) \leq \varepsilon_1 + \varepsilon_2.$$

3 Mirror Theory

The goal of this section is to lower bound the number of solutions to a certain type of systems of equations and non-equations. We will represent a system of equations and non-equations by a simple graph containing no loops or multiple edges; each vertex denotes an n-bit unknown (for a fixed n), and each edge is labeled with an element in $\{0,1\}^n \cup \{\neq\}$, where $\neq$ is a special symbol meaning non-equality. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph and let $\overline{PQ} \in \mathcal{E}$ be an edge for $P, Q \in \mathcal{V}$. If this edge is labeled with $\lambda \in \{0,1\}^n$, then it means an equation $P \oplus Q = \lambda$, while if it is labeled with a special symbol $\neq$, then it means that P and Q are distinct. We will sometimes write $P \overset{\star}{-} Q$ when an edge $\overline{PQ}$ is labeled with $\star \in \{0,1\}^n \cup \{\neq\}$.

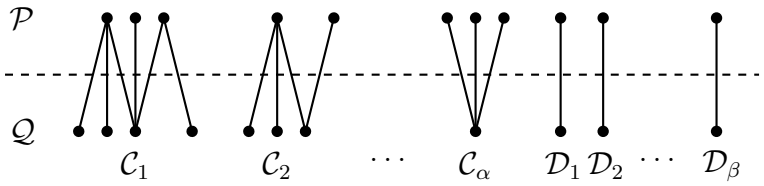


Fig. 2: A bipartite graph $\mathcal{G}^{=}$ with two parts $\mathcal{P}$ and $\mathcal{Q}$.

Let $\mathcal{G}^{=}$ denote the graph obtained by deleting all $\neq$-labeled edges from $\mathcal{G}$. For $\ell > 0$ and a trail[1]

$$\mathcal{L} : P_0 \overset{\lambda_1}{-} P_1 \overset{\lambda_2}{-} \cdots \overset{\lambda_\ell}{-} P_\ell$$

in $\mathcal{G}^{=}$, its label is defined as

$$\lambda(\mathcal{L}) \overset{\text{def}}{=} \lambda_1 \oplus \lambda_2 \oplus \cdots \oplus \lambda_\ell.$$

In this work, we will focus on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with certain properties, as listed below.

1. $\mathcal{G}^{=}$ contains no isolated vertex; every vertex is incident with at least one edge.
2. The vertex set $\mathcal{V}$ is partitioned into two disjoint parts, denoted $\mathcal{P}$ and $\mathcal{Q}$; the edge set $\mathcal{E}$ contains $P \overset{\neq}{-} P'$ for any different P, $P' \in \mathcal{P}$, and $Q \overset{\neq}{-} Q'$ for any different Q, $Q' \in \mathcal{Q}$.
3. $\mathcal{G}^{=}$ contains no cycle.
4. $\lambda(\mathcal{L}) \neq \mathbf{0}$ for any trail $\mathcal{L}$ of even length in $\mathcal{G}^{=}$.

Any graph $\mathcal{G}$ satisfying the above properties will be called a *nice* graph. For a nice graph $\mathcal{G}$, $\mathcal{G}^{=}$ is a bipartite graph with no cycle, where every edge connects a vertex in $\mathcal{P}$ to one in $\mathcal{Q}$. So $\mathcal{G}^{=}$ is decomposed into its connected components, all of which are trees; let

$$\mathcal{G}^{=} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \cdots \sqcup \mathcal{C}_\alpha \sqcup \mathcal{D}_1 \sqcup \mathcal{D}_2 \sqcup \cdots \sqcup \mathcal{D}_\beta$$

for some α, $\beta \geq 0$, where $\mathcal{C}_i$ denotes a component of size greater than 2, and $\mathcal{D}_i$ denotes a component of size 2. We will also write $\mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \cdots \sqcup \mathcal{C}_\alpha$ and $\mathcal{D} = \mathcal{D}_1 \sqcup \mathcal{D}_2 \sqcup \cdots \sqcup \mathcal{D}_\beta$ (Figure 2).

Any solution to $\mathcal{G}$ (identifying $\mathcal{G}$ with its corresponding system of equations and non-equations) should satisfy all the equations in $\mathcal{G}^{=}$, while all the variables in $\mathcal{P}$ (resp. $\mathcal{Q}$) should take on different values. We remark that if we assign any value to a vertex P, then the labeled edges determine the values of all the other vertices in the component containing P, where the assignment is unique since $\mathcal{G}^{=}$

[1] A trail is a walk in which all edges are distinct.

contains no cycle, and the values in the same part are all distinct since $\lambda(\mathcal{L}) \neq \mathbf{0}$ for any trail $\mathcal{L}$ of even length.

On the other hand, the number of possible assignments of distinct values to the vertices in $\mathcal{P}$ (resp. $\mathcal{Q}$) is $(N)_{|\mathcal{P}|}$ (resp. $(N)_{|\mathcal{Q}|}$). One might expect that when such an assignment is chosen uniformly at random, it would satisfy all the equations in $\mathcal{G}^{=}$ with probability $1/N^q$, where q denotes the number of edges (i.e., equations) in $\mathcal{G}^{=}$. Indeed, we can prove that the number of solutions to $\mathcal{G}$ is close to $\frac{(N)_{|\mathcal{P}|}(N)_{|\mathcal{Q}|}}{N^q}$ up to a certain error (that can be negligible according to the parameters).

Theorem 1. *Let $\mathcal{G}$ be a nice graph, and let q and q_c denote the number of edges of $\mathcal{G}^{=}$ and $\mathcal{C}$, respectively. If $q < \frac{N}{8}$, then the number of solutions to $\mathcal{G}$, denoted $h(\mathcal{G})$, satisfies*

$$\frac{h(\mathcal{G})N^q}{(N)_{|\mathcal{P}|}(N)_{|\mathcal{Q}|}} \geq 1 - \frac{9q_c^2}{8N} - \frac{3q_cq^2}{2N^2} - \frac{q^2}{N^2} - \frac{9q_c^2q}{8N^2} - \frac{8q^4}{3N^3}.$$

Proof. For $i = 1, \ldots, \alpha$, $\mathcal{C}_i$ is a bipartite graph, where one part consists of the vertices in $\mathcal{P}$ and the other vertices in $\mathcal{Q}$; the two parts are denoted $\mathcal{P}_i$ and $\mathcal{Q}_i$, respectively. Let $r_i = |\mathcal{P}_i|$ and $s_i = |\mathcal{Q}_i|$, let $d_i = r_i + s_i$.

Let $h_c(i)$ be the number of solutions to $\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$. In order to find a relation between $h_c(i)$ and $h_c(i+1)$, we fix a solution to $\mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_i$. If we fix a vertex $P^* \in \mathcal{P}_{i+1}$ and assign any value to P^*, then the other unknowns are uniquely determined, since there is a unique trail from P^* to any other vertex in $\mathcal{C}_{i+1}$. In order to satisfy the non-equations, it is sufficient that

$$P^* \notin \bigcup_{\substack{1 \leq j \leq i \\ P \in \mathcal{P}_{i+1}}} (\mathcal{P}_j \oplus \lambda_P) \cup \bigcup_{\substack{1 \leq j \leq i \\ Q \in \mathcal{Q}_{i+1}}} (\mathcal{Q}_j \oplus \lambda_Q),$$

where λ_X denotes the label of the unique trail from P^* to X if $X \neq P^*$ and $\lambda_{P^*} = \mathbf{0}$. The number of such choices is at least

$$N - (r_1 + \cdots + r_i)r_{i+1} - (s_1 + \cdots + s_i)s_{i+1}.$$

Then we have

$$\begin{aligned}
h_c(\alpha) &\geq N^\alpha \left(1 - \frac{r_1r_2 + s_1s_2}{N}\right) \cdots \left(1 - \frac{(r_1 + \cdots + r_{\alpha-1})r_\alpha + (s_1 + \cdots + s_{\alpha-1})s_\alpha}{N}\right) \\
&\geq N^\alpha \left(1 - \frac{1}{N} \sum_{1 \leq i < j \leq \alpha} (r_ir_j + s_is_j)\right) \\
&\geq N^\alpha \left(1 - \frac{1}{2N} \left(\sum_{i=1}^{\alpha} d_i\right)^2\right) \\
&\geq N^\alpha \left(1 - \frac{9q_c^2}{8N}\right) \qquad (1)
\end{aligned}$$

since $h_c(1) = N$, $\sum_{i=1}^{\alpha} d_i = \alpha + q_c$ and $\alpha \leq q_c/2$.

For $i = 1, \ldots, \beta$, we will write

$$\mathcal{D}_i : P_i \overset{\lambda_i}{-} Q_i$$

where $P_i \in \mathcal{P}$ and $Q_i \in \mathcal{Q}$. Let $h_d(i)$ be the number of solutions to $\mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i$ for $i = 1, \ldots, \beta$. Note that $h_d(0) = h_c(\alpha)$ and $h_d(\beta) = h(\mathcal{G})$. In order to find a relation between $h_d(i)$ and $h_d(i+1)$, we fix a solution to $\mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i$. Then we can choose P_{i+1} from $\{0,1\}^n \setminus (\mathcal{X}_i \cup (\mathcal{Y}_i \oplus \lambda_{i+1}))$, where

$$\mathcal{X}_i \overset{\text{def}}{=} \bigsqcup_{1 \leq j \leq \alpha} \mathcal{P}_j \sqcup \{P_1, \ldots, P_i\},$$
$$\mathcal{Y}_i \overset{\text{def}}{=} \bigsqcup_{1 \leq j \leq \alpha} \mathcal{Q}_j \sqcup \{Q_1, \ldots, Q_i\}.$$

For $i = 0, \ldots, \beta - 1$, let

$$R_i = r_1 + \cdots + r_\alpha + i,$$
$$S_i = s_1 + \cdots + s_\alpha + i.$$

Then, since $|\mathcal{X}_i| = R_i$ and $|\mathcal{Y}_i| = S_i$, we have

$$\begin{aligned} h_d(i+1) &= \sum_{\substack{\text{solutions to} \\ \mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i}} (N - |\mathcal{X}_i \cup (\mathcal{Y}_i \oplus \lambda_{i+1})|) \\ &= \sum_{\substack{\text{solutions to} \\ \mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i}} (N - R_i - S_i + |\mathcal{X}_i \cap (\mathcal{Y}_i \oplus \lambda_{i+1})|) \\ &= (N - R_i - S_i) h_d(i) + \sum_{\substack{\text{solutions to} \\ \mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i}} |\mathcal{X}_i \cap (\mathcal{Y}_i \oplus \lambda_{i+1})|. \end{aligned} \tag{2}$$

For $X \in \mathcal{X}_i$ and $Y \in \mathcal{Y}_i$, let $h'(X, Y)$ denote the number of solutions to $\mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i$ such that $X \oplus Y = \lambda_{i+1}$. Then we have

$$\begin{aligned} \sum_{\substack{\text{solutions to} \\ \mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i}} |\mathcal{X}_i \cap (\mathcal{Y}_i \oplus \lambda_{i+1})| &= \sum_{X \in \mathcal{X}_i, Y \in \mathcal{Y}_i} h'(X, Y) \\ &\geq \sum_{\substack{X \in \{P_1, \ldots, P_i\} \\ Y \in \{Q_1, \ldots, Q_i\}}} h'(X, Y). \end{aligned} \tag{3}$$

If $X = P_j$, $Y = Q_j$, and $\lambda_{i+1} = \lambda_j$ for some $j = 1, \ldots, i$, then the additional equation $X \oplus Y = \lambda_{i+1}$ is redundant, and hence $h'(X, Y) = h_d(i)$. Suppose that $X = P_j$ and $Y = Q_{j'}$ for distinct j and j', and $\lambda_{i+1} \notin \{\lambda_j, \lambda_{j'}\}$. In this case, and for $i \geq 2$, we have

$$h'(X, Y) \geq \frac{h_d(i)}{N} \left(1 - \frac{2(R_i + S_i)}{N}\right) \tag{4}$$

since

$$h'(X,Y) \geq (N - 2(R_i + S_i - 4))\, h_d(i-2) \geq (N - 2(R_i + S_i))\, h_d(i-2),$$
$$h_d(i-2)N^2 \geq h_d(i-2)\,(N - (R_i + S_i - 4))\,(N - (R_i + S_i - 2)) \geq h_d(i).$$

Let

$$G = \left|\{1 \leq j \leq i : \lambda_j = \lambda_{i+1}\}\right|,$$
$$H = \left|\left\{(j,j') \in [i]^{*2} : \lambda_j \neq \lambda_{i+1}, \lambda_{j'} \neq \lambda_{i+1}\right\}\right|.$$

Then we have

$$H \geq i(i-1) - 2iG. \tag{5}$$

By (3), (4), (5), and since $2i \leq 2q \leq N$, we have

$$\begin{aligned}\sum_{\substack{\text{solutions to} \\ \mathcal{C} \sqcup \mathcal{D}_1 \sqcup \cdots \sqcup \mathcal{D}_i}} |\mathcal{X}_i \cap (\mathcal{Y}_i \oplus \lambda_{i+1})| &\geq \left(G + \frac{i(i-1) - 2iG}{N}\left(1 - \frac{2(R_i + S_i)}{N}\right)\right) h_d(i) \\ &\geq \frac{i(i-1)}{N}\left(1 - \frac{2(R_i + S_i)}{N}\right) h_d(i),\end{aligned}$$

and by (2),

$$h_d(i+1) \geq (N - R_i - S_i) h_d(i) + \frac{i(i-1)}{N}\left(1 - \frac{2(R_i + S_i)}{N}\right) h_d(i).$$

Since $\frac{R_i + S_i}{2} \leq q < \frac{N}{8}$ and $R_0 + S_0 = \alpha + q_c \leq \frac{3q_c}{2}$, we have

$$\begin{aligned}\frac{h_d(i+1)N}{h_d(i)(N - R_i)(N - S_i)} &\geq \frac{N^2 - (R_i + S_i)N + (i^2 - i)\left(1 - \frac{2(R_i + S_i)}{N}\right)}{N^2 - (R_i + S_i)N + R_i S_i} \\ &= 1 - \frac{R_i S_i - (i^2 - i)\left(1 - \frac{2(R_i + S_i)}{N}\right)}{N^2 - (R_i + S_i)N + R_i S_i} \\ &\geq 1 - \frac{(R_0 + i)(S_0 + i) - (i^2 - i) + \frac{2(R_i + S_i)i^2}{N}}{N^2/2} \\ &\geq 1 - \frac{2R_0 S_0}{N^2} - \frac{2(R_0 + S_0 + 1)i}{N^2} - \frac{4(R_i + S_i)i^2}{N^3} \\ &\geq 1 - \frac{9q_c^2}{8N^2} - \frac{3q_c i + 2i}{N^2} - \frac{8qi^2}{N^3}.\end{aligned} \tag{6}$$

Since $q = q_c + \beta$, $|\mathcal{P}| = R_0 + \beta$, $|\mathcal{Q}| = S_0 + \beta$ and $\alpha + q_c = R_0 + S_0$, and by (1) and (6), we have

$$\begin{aligned}
\frac{h(\mathcal{G})N^q}{(N)_{|\mathcal{P}|}(N)_{|\mathcal{Q}|}} &= \frac{h(\mathcal{G})N^{q_c+\beta}}{(N)_{R_0}(N-R_0)_\beta(N)_{S_0}(N-S_0)_\beta} \\
&= \frac{h_c(\alpha)N^{q_c}}{(N)_{R_0}(N)_{S_0}} \prod_{i=0}^{\beta-1} \left(\frac{h_d(i+1)N}{h_d(i)(N-R_i)(N-S_i)} \right) \\
&\geq \frac{h_c(\alpha)}{N^\alpha} \prod_{i=0}^{\beta-1} \left(\frac{h_d(i+1)N}{h_d(i)(N-R_i)(N-S_i)} \right) \\
&\geq \left(1 - \frac{9q_c^2}{8N}\right) \prod_{i=0}^{\beta-1} \left(1 - \frac{9q_c^2}{8N^2} - \frac{3q_c i + 2i}{N^2} - \frac{8qi^2}{N^3}\right) \\
&\geq \left(1 - \frac{9q_c^2}{8N}\right) \left(1 - \sum_{i=0}^{\beta-1} \left(\frac{9q_c^2}{8N^2} + \frac{3q_c i + 2i}{N^2} + \frac{8qi^2}{N^3}\right)\right) \\
&\geq \left(1 - \frac{9q_c^2}{8N}\right) \left(1 - \frac{9q_c^2 q}{8N^2} - \frac{3q_c q^2}{2N^2} - \frac{q^2}{N^2} - \frac{8q^4}{3N^3}\right) \\
&\geq 1 - \frac{9q_c^2}{8N} - \frac{9q_c^2 q}{8N^2} - \frac{3q_c q^2}{2N^2} - \frac{q^2}{N^2} - \frac{8q^4}{3N^3}
\end{aligned}$$

which completes the proof. □

4 A Framework for Security Proof of DbHtS MACs

In this section, we consider $\mathsf{DbHtS}[H, E]$ based on a $2n$-bit function H and an n-bit block cipher E. A message M is encrypted as

$$E_{K_1}(F_{K_h}(M)) \oplus E_{K_2}(G_{K_h}(M))$$

by keys K_h, K_1 and K_2, where we write $H_{K_h}(M) = (F_{K_h}(M), G_{K_h}(M))$ (see Section 2).

Up to the PRP-security of E, the keyed permutations E_{K_1} and E_{K_2} can be replaced by independent random permutations π_1 and π_2, in which case we simply write $\mathsf{DbHtS}[H]$ instead of $\mathsf{DbHtS}[H, E]$. The goal of this section is to establish a general framework for security proof of $\mathsf{DbHtS}[H]$ using Theorem 1.

Graph Representation of Transcripts. At the end of the attack, the distinguisher $\mathcal{D}$ will be given K_h for free. Then, from the transcript

$$\tau = (K_h, (M_i, T_i)_{1 \leq i \leq q}),$$

$H_{K_h}(M_i) = (U_i, V_i)$ are fixed for $i = 1, \ldots, q$. The core of the security proof is to estimate the number of possible ways of fixing $\pi_1(U_i)$ and $\pi_2(V_i)$ in a way that

$\pi_1(U_i) \oplus \pi_2(V_i) = T_i$ for $i = 1, \ldots, q$. So $\{\pi_1(U_i)\}$ and $\{\pi_2(V_i)\}$ are identified with two sets of unknowns

$$\mathcal{P} = \{P_1, \ldots, P_{q_1}\},$$
$$\mathcal{Q} = \{Q_1, \ldots, Q_{q_2}\},$$

respectively, where q_1, $q_2 \leq q$, since there might be collisions between U_i's or between V_i's. Assuming that $\mathcal{P}$ and $\mathcal{Q}$ are disjoint, we connect P_j and $Q_{j'}$ with an edge of label T_i if $\pi_1(U_i) = P_j$ and $\pi_2(V_i) = Q_{j'}$ for some i. Any pair of vertices in the same set of either $\mathcal{P}$ or $\mathcal{Q}$ are connected by a $\neq$-labeled edge. In this way, we obtain a graph on $\mathcal{P} \sqcup \mathcal{Q}$, called the *transcript graph* of τ and denoted $\mathcal{G}_\tau$.

Good Transcripts. Fix a parameter $\bar{q}_c$ (to be optimized later). A transcript $\tau = (K_h, (M_i, T_i)_{1 \leq i \leq q})$ is defined as *good* if

1. the transcript graph $\mathcal{G}_\tau$ is nice (as defined in Section 3);
2. the number of edges in $\mathcal{C}$ (i.e., edges in the components of size greater than two) is not greater than $\bar{q}_c$.

If a transcript τ is not good, then it will be called a *bad* transcript.

For a transcript graph $\mathcal{G}_\tau$, let $\mathcal{G}_\tau^=$ denote the graph obtained by deleting all $\neq$-labeled edges from $\mathcal{G}_\tau$. Then $\mathcal{G}_\tau^=$ is a bipartite graph with q edges. By definition, $\mathcal{G}_\tau^=$ has no isolated vertices. So in order to see if $\mathcal{G}_\tau$ is nice, it is sufficient to check out if

1. $\mathcal{G}_\tau^=$ has no cycle;
2. $\lambda(\mathcal{L}) \neq \mathbf{0}$ for any trail $\mathcal{L}$ of even length.

A Framework for Security Proof. Once bad transcripts have been defined, we will show that

$$\Pr[\mathsf{T}_{\mathrm{id}} \in \mathit{\Gamma}_{\mathsf{bad}}] \leq \varepsilon_2$$

for a small $\varepsilon_2 > 0$. Next, we fix a good transcript τ. Obviously, we have

$$\Pr[\mathsf{T}_{\mathrm{id}} = \tau] = \frac{1}{|\mathcal{K}_h| \cdot N^q}.$$

The probability of obtaining τ in the real world is computed over the randomness of π_1 and π_2. By Theorem 1 and since $q_c \leq \bar{q}_c$, the number of possible ways of fixing $\pi_1(U_i)$ and $\pi_2(V_i)$ (i.e., $h(\mathcal{G}_\tau)$) is lower bounded by

$$\frac{(N)_{|\mathcal{P}|}(N)_{|\mathcal{Q}|}}{N^q}(1 - \varepsilon_1)$$

where

$$\varepsilon_1 \stackrel{\mathrm{def}}{=} \frac{9\bar{q}_c^2}{8N} + \frac{3\bar{q}_c q^2}{2N^2} + \frac{q^2}{N^2} + \frac{9\bar{q}_c^2 q}{8N^2} + \frac{8q^4}{3N^3}. \tag{7}$$

The probability that π_1 and π_2 realize each assignment is exactly $1/(N)_{|\mathcal{P}|}(N)_{|\mathcal{Q}|}$. So we have

$$\frac{\Pr\left[\mathsf{T}_{\mathrm{re}} = \tau\right]}{\Pr\left[\mathsf{T}_{\mathrm{id}} = \tau\right]} \geq 1 - \varepsilon_1,$$

and by Lemma 1,

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{DbHtS}[H]}(\mathcal{D}) \leq \varepsilon_1 + \varepsilon_2.$$

5 Concatenating Universal Hash Functions

In this section, we will prove the security of DbHtS when the underlying hash function H is defined as the concatenation of two copies of an almost universal hash function using independent keys.

Let $\delta > 0$, and let $F : \mathcal{K} \times \mathcal{M} \to \{0,1\}^n$ be a δ-almost universal hash function. We will consider $\mathsf{DbHtS}[H]$, where

$$\begin{aligned} H : (\mathcal{K} \times \mathcal{K}) \times \mathcal{M} &\longrightarrow \{0,1\}^n \times \{0,1\}^n \\ ((K_1, K_2), M) &\longmapsto (F_{K_1}(M), F_{K_2}(M)). \end{aligned}$$

We fix the parameter $\bar{q}_c$, and define bad events as follows.

- $\mathsf{Bad}_1 \Leftrightarrow$ there is a pair of distinct queries (M_i, M_j) such that $F_{K_1}(M_i) = F_{K_1}(M_j)$ and $F_{K_2}(M_i) = F_{K_2}(M_j)$.
- $\mathsf{Bad}_2 \Leftrightarrow \mathsf{Bad}_{2a} \vee \mathsf{Bad}_{2b}$, where
  - $\mathsf{Bad}_{2a} \Leftrightarrow$ there is a quadruple of distinct queries $(M_{i_1}, M_{i_2}, M_{i_3}, M_{i_4})$ such that $F_{K_1}(M_{i_1}) = F_{K_1}(M_{i_2})$, $F_{K_2}(M_{i_2}) = F_{K_2}(M_{i_3})$, $F_{K_1}(M_{i_3}) = F_{K_1}(M_{i_4})$,
  - $\mathsf{Bad}_{2b} \Leftrightarrow$ there is a quadruple of distinct queries $(M_{i_1}, M_{i_2}, M_{i_3}, M_{i_4})$ such that $F_{K_2}(M_{i_1}) = F_{K_2}(M_{i_2})$, $F_{K_1}(M_{i_2}) = F_{K_1}(M_{i_3})$, $F_{K_2}(M_{i_3}) = F_{K_2}(M_{i_4})$.
- $\mathsf{Bad}_3 \Leftrightarrow$ there is a pair of distinct queries (M_i, M_j) such that $T_i \oplus T_j = \mathbf{0}$ and either $F_{K_1}(M_i) = F_{K_1}(M_j)$ or $F_{K_2}(M_i) = F_{K_2}(M_j)$.
- $\mathsf{Bad}_4 \Leftrightarrow \mathsf{Bad}_{4a} \vee \mathsf{Bad}_{4b}$, where
  - $\mathsf{Bad}_{4a} \Leftrightarrow$ the number of distinct queries (M_i, M_j) such that $F_{K_1}(M_i) = F_{K_1}(M_j)$ is greater than $\bar{q}_c/4$,
  - $\mathsf{Bad}_{4b} \Leftrightarrow$ the number of distinct queries (M_i, M_j) such that $F_{K_2}(M_i) = F_{K_2}(M_j)$ is greater than $\bar{q}_c/4$.

We observe that

1. $\mathcal{G}^{=}_{\tau}$ contains no cycle of length 2 without Bad_1;
2. $\mathcal{G}^{=}_{\tau}$ contains no trail of length 4 without Bad_2;
3. $\lambda(\mathcal{L}) \neq \mathbf{0}$ for any trail $\mathcal{L}$ of length 2 without Bad_3.

A distinct pair of "half-colliding" queries such that either $F_{K_1}(M_i) = F_{K_1}(M_j)$ or $F_{K_2}(M_i) = F_{K_2}(M_j)$ will add an edge to any component containing it, and make the size of the component greater than two; the number of edges in $\mathcal{C}$ cannot be twice as many as the number of half-collisions. So the number of edges in $\mathcal{C}$ is not greater than $\bar{q}_c$ without Bad_4. With this observation, we conclude that a transcript is good without any bad event above; namely,

$$\Pr[\mathsf{T}_{\mathrm{id}} \in \mathit{\Gamma}_{\mathsf{bad}}] \leq \Pr[\mathsf{Bad}_1 \vee \mathsf{Bad}_2 \vee \mathsf{Bad}_3 \vee \mathsf{Bad}_4].$$

We can upper bound the probability of each bad event as follows.

1. The probability that there exists a pair of distinct queries (M_i, M_j) such that $F_{K_1}(M_i) = F_{K_1}(M_j)$ and $F_{K_2}(M_i) = F_{K_2}(M_j)$ is upper bounded by $q^2\delta^2$ since K_1 and K_2 are independent. Namely,

$$\Pr[\mathsf{Bad}_1] \leq q^2\delta^2.$$

2. By the Markov inequality, we have

$$\Pr[\mathsf{Bad}_{4a}],\ \Pr[\mathsf{Bad}_{4b}] \leq \frac{4q^2\delta}{\bar{q}_c}.$$

3. Given that the number of F_{K_1}-collisions is upper bounded by $\bar{q}_c/4$, the probability that there exist two F_{K_1}-colliding pairs (M_{i_1}, M_{i_2}) and (M_{i_3}, M_{i_4}) such that $F_{K_2}(M_{i_2}) = F_{K_2}(M_{i_3})$ is upper bounded by $\frac{\bar{q}_c^2\delta}{16}$. Namely, we have

$$\Pr[\mathsf{Bad}_{2a} \mid \neg\mathsf{Bad}_{4a}] \leq \frac{\bar{q}_c^2\delta}{16}.$$

   Similarly, we have

$$\Pr[\mathsf{Bad}_{2b} \mid \neg\mathsf{Bad}_{4b}] \leq \frac{\bar{q}_c^2\delta}{16}.$$

4. For each pair of distinct queries (M_i, M_j), the probability that $T_i \oplus T_j = \mathbf{0}$ is $1/N$, and the probability that either $F_{K_1}(M_i) = F_{K_1}(M_j)$ or $F_{K_2}(M_i) = F_{K_2}(M_j)$ is upper bounded by δ. Since the two events are independent and by the union bound, we have

$$\Pr[\mathsf{Bad}_3] \leq \frac{q^2\delta}{N}.$$

All in all, we have

$$\begin{aligned}
\Pr[\mathsf{T}_{\mathrm{id}} \in \mathit{\Gamma}_{\mathsf{bad}}] &\leq \Pr[\mathsf{Bad}_1 \vee \mathsf{Bad}_2 \vee \mathsf{Bad}_3 \vee \mathsf{Bad}_4] \\
&\leq \Pr[\mathsf{Bad}_1] + \Pr[\mathsf{Bad}_3] + \Pr[\mathsf{Bad}_{4a}] + \Pr[\mathsf{Bad}_{2a} \mid \neg\mathsf{Bad}_{4a}] \\
&\quad + \Pr[\mathsf{Bad}_{4b}] + \Pr[\mathsf{Bad}_{2b} \mid \neg\mathsf{Bad}_{4b}] \\
&\leq q^2\delta^2 + \frac{q^2\delta}{N} + \frac{8q^2\delta}{\bar{q}_c} + \frac{\bar{q}_c^2\delta}{8}. \qquad (8)
\end{aligned}$$

Combining (7) and (8), we have

$$\begin{aligned}\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{DbHtS}[H]}(\mathcal{D}) \le q^2\delta^2 + \frac{q^2\delta}{N} + \frac{8q^2\delta}{\bar{q}_c} + \frac{\bar{q}_c^2\delta}{8} \\ + \frac{9\bar{q}_c^2}{8N} + \frac{3\bar{q}_c q^2}{2N^2} + \frac{q^2}{N^2} + \frac{9\bar{q}_c^2 q}{8N^2} + \frac{8q^4}{3N^3}\end{aligned}$$

for any distinguisher $\mathcal{D}$ making q queries, and for any $\bar{q}_c > 0$. When $\bar{q}_c = 4q^{\frac{2}{3}}$ (by setting $8q^2\delta/\bar{q}_c = \bar{q}_c^2\delta/8$), we obtain the following theorem.

Theorem 2. *Let $\delta > 0$, and let $F : \mathcal{K} \times \mathcal{M} \to \{0,1\}^n$ be a δ-almost universal hash function. Let*

$$\begin{aligned}H : (\mathcal{K} \times \mathcal{K}) \times \mathcal{M} &\longrightarrow \{0,1\}^n \times \{0,1\}^n \\ ((K_1, K_2), M) &\longmapsto (F_{K_1}(M), F_{K_2}(M)).\end{aligned}$$

Then one has

$$\begin{aligned}\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{DbHtS}[H]}(\mathcal{D}) \le 4q^{\frac{4}{3}}\delta + q^2\delta^2 + \frac{q^2\delta}{N} + \frac{18q^{\frac{4}{3}}}{N} \\ + \frac{6q^{\frac{8}{3}}}{N^2} + \frac{18q^{\frac{7}{3}}}{N^2} + \frac{q^2}{N^2} + \frac{8q^4}{3N^3}.\end{aligned}$$

When $\delta \approx \frac{1}{N}$, $\mathsf{DbHtS}[H]$ becomes a PRF that is secure up to $2^{\frac{3n}{4}}$ queries.

5.1 Security of PolyMAC

An n-bit keyed function PolyHash is defined with key space $\mathcal{K} = \{0,1\}^n$, where $\{0,1\}^n$ is identified with a finite field $\mathbf{GF}(2^n)$ with 2^n elements. For a padded message $M = M[1]\|M[2]\|\cdots\|M[m]$ where $m \le \ell$, and a key $K \in \mathcal{K}$, $\mathsf{PolyHash}_K(M)$ is defined using finite field addition and multiplication, denoted $\oplus$ and $\cdot$, respectively.

Function $\mathsf{PolyHash}_K(M)$
  $Z[0] \leftarrow 0$
  for $\alpha \leftarrow 1$ to m **do**
    $Z[\alpha] \leftarrow K \cdot (Z[\alpha - 1] \oplus M[\alpha])$
  return $Z[m]$

The PolyMAC MAC is defined as $\mathsf{DbHtS}[H]$, where

$$\begin{aligned}H : (\mathcal{K} \times \mathcal{K}) \times \mathcal{M} &\longrightarrow \{0,1\}^n \times \{0,1\}^n \\ ((K_1, K_2), M) &\longmapsto (\mathsf{PolyHash}_{K_1}(M), \mathsf{PolyHash}_{K_2}(M)).\end{aligned}$$

It is not hard to show that PolyHash is $\frac{\ell}{N}$-almost universal. Therefore, by Theorem 2, we obtain the following theorem.

Theorem 3. *When* PolyMAC *is based on a block cipher E, one has*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{PolyMAC}}(q,t,\ell) \le \frac{(4\ell+18)q^{\frac{4}{3}}}{N} + \frac{6q^{\frac{8}{3}}}{N^2} + \frac{18q^{\frac{7}{3}}}{N^2} + \frac{(\ell^2+\ell+1)q^2}{N^2} + \frac{8q^4}{3N^3} + 2\mathbf{Adv}^{\mathsf{prp}}_{E}(q,t+t'),$$

where t' is the time complexity necessary to compute E for q times.

5.2 Security of SUM-ECBC

An n-bit hash function CBC is based an n-bit block cipher E using k-bit keys. For a padded message $M = M[1]\|M[2]\|\cdots\|M[m]$ where $m \le \ell$, and a key $K \in \{0,1\}^k$, $\mathsf{CBC}_K(M)$ is defined as follows.

Function $\mathsf{CBC}_K(M)$
  $Z[0] \leftarrow 0$
  for $\alpha \leftarrow 1$ to m **do**
    $Z[\alpha] \leftarrow E_K\left(Z[\alpha-1] \oplus M[\alpha]\right)$
  return $Z[m]$

The SUM-ECBC MAC is defined as $\mathsf{DbHtS}[H]$ (Figure 3), where

$$H : (\{0,1\}^k \times \{0,1\}^k) \times \mathcal{M} \longrightarrow \{0,1\}^n \times \{0,1\}^n$$
$$((K_1,K_2),M) \longmapsto (\mathsf{CBC}_{K_1}(M), \mathsf{CBC}_{K_2}(M)).$$

For $m \le \ell$, let $d(m)$ be the number of divisors of m and let $d'(\ell) = \max_{m\le\ell} d(m)$. It is known that $d'(\ell) = \ell^{o(1)}$. In [11, Corollary 2], it has been proved that CBC is δ-almost universal when the underlying block cipher is replaced by a truly random permutation, where

$$\delta = \frac{d'(\ell)}{N-2\ell} + \frac{16\ell^4}{N^2}.$$

Therefore, by Theorem 2, we obtain the following theorem.

Theorem 4. *Assume that $\ell \le N/4$. When* SUM-ECBC *is based on a block cipher E, one has*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{SUM\text{-}ECBC}}(q,t,\ell) \le \frac{(8d'(\ell)+18)q^{\frac{4}{3}}}{N} + \frac{6q^{\frac{8}{3}}}{N^2} + \frac{18q^{\frac{7}{3}}}{N^2} + \frac{(4d'(\ell)^2+2d'(\ell)+1)q^2}{N^2} + \frac{64\ell^4 q^{\frac{4}{3}}}{N^2} + \frac{8q^4}{3N^3} + \frac{(64d'(\ell)+16)\ell^4q^2}{N^3} + \frac{256\ell^8q^2}{N^4} + 4\mathbf{Adv}^{\mathsf{prp}}_{E}(\ell q, t+t'),$$

where t' is the time complexity necessary to compute E for ℓq times.

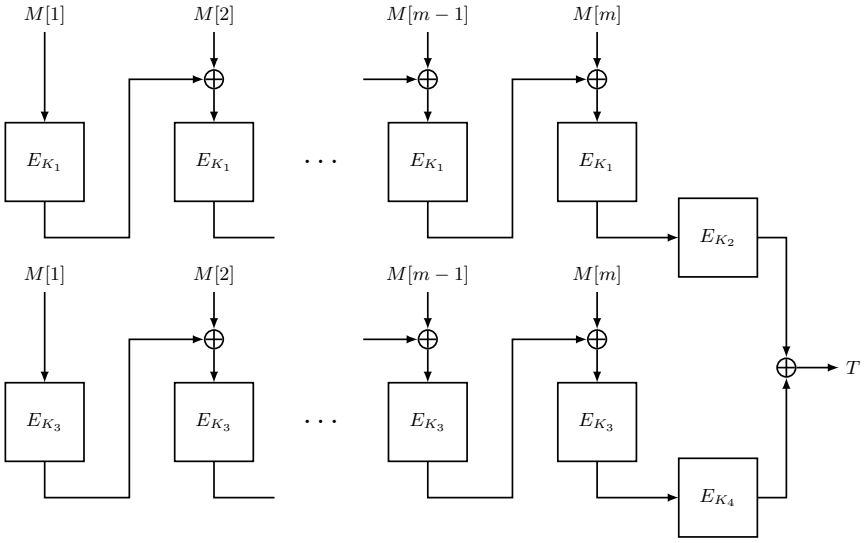


Fig. 3: SUM-ECBC based on a block cipher E using four keys K_i, $i = 1, 2, 3, 4$.

6 Security of PMAC-Plus

A $2n$-bit hash function PHash is based an n-bit block cipher E using k-bit keys. For a padded message $M = M[1]\|M[2]\|\cdots\|M[m]$ where $m \leq \ell$, and a key $K \in \{0,1\}^k$, $\mathsf{PHash}_K(M)$ is defined as follows.

Function $\mathsf{PHash}_K(M)$
  $\Delta_0 \leftarrow E_K(0)$
  $\Delta_1 \leftarrow E_K(1)$
  for $\alpha \leftarrow 1$ to m **do**
    $X[\alpha] \leftarrow M[\alpha] \oplus 2^{\alpha} \cdot \Delta_0 \oplus 2^{2\alpha} \cdot \Delta_1$
    $Y[\alpha] \leftarrow E_K(X[\alpha])$
  $U \leftarrow Y[1] \oplus Y[2] \oplus \cdots \oplus Y[m]$
  $V \leftarrow Y[1] \oplus 2 \cdot Y[2] \oplus \cdots \oplus 2^{m-1} \cdot Y[m]$
  return (U, V)

The PMAC-Plus MAC is defined as DbHtS[PHash] (Figure 4).

For simplicity of proof, we will replace keyed permutations E_{K_1}, E_{K_2}, E_{K_3} by independent random permutations π, π', π'', respectively, up to the PRP-security of E (to be captured by the term $3\mathbf{Adv}^{\mathsf{prp}}_E(\ell q, t + t')$ in Theorem 5). So we will focus on PHash based on a truly random permutation π, and upper bound the probability of bad transcripts (as defined in Section 4).[2]

[2] We will simply omit key $\pi \in \mathsf{Perm}(n)$ in PHash and its halves F and G.

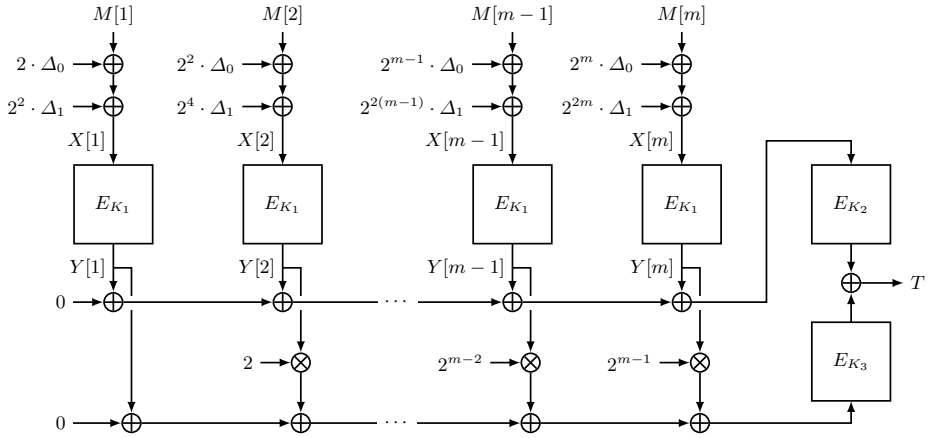


Fig. 4: PMAC-Plus based on a block cipher E using three keys K_1, K_2, K_3, where $\Delta_0 = E_{K_1}(0)$ and $\Delta_1 = E_{K_1}(1)$.

Bad Events. Note that $\mathsf{PHash}(M) = (F(M), G(M))$ for any message M. We fix a parameter $\bar{q}_c$, and define bad events as follows.

- $\mathsf{Bad}_1 \Leftrightarrow$ there is a pair of distinct queries (M_i, M_j) such that $\mathsf{PHash}(M_i) = \mathsf{PHash}(M_j)$.
- $\mathsf{Bad}_2 \Leftrightarrow$ there is a quadruple of distinct queries $(M_{i_1}, M_{i_2}, M_{i_3}, M_{i_4})$ such that $F(M_{i_1}) = F(M_{i_2})$, $G(M_{i_2}) = G(M_{i_3})$, $F(M_{i_3}) = F(M_{i_4})$.
- $\mathsf{Bad}_3 \Leftrightarrow$ there is a quadruple of distinct queries $(M_{i_1}, M_{i_2}, M_{i_3}, M_{i_4})$ such that $G(M_{i_1}) = G(M_{i_2})$, $F(M_{i_2}) = F(M_{i_3})$, $G(M_{i_3}) = G(M_{i_4})$ and $T_{i_1} \oplus T_{i_2} \oplus T_{i_3} \oplus T_{i_4} = \mathbf{0}$.
- $\mathsf{Bad}_4 \Leftrightarrow$ there is a pair of distinct queries (M_i, M_j) such that $T_i \oplus T_j = \mathbf{0}$ and either $F(M_i) = F(M_j)$ or $G(M_i) = G(M_j)$.
- $\mathsf{Bad}_5 \Leftrightarrow \mathsf{Bad}_{5a} \vee \mathsf{Bad}_{5b}$, where
  - $\mathsf{Bad}_{5a} \Leftrightarrow$ the number of distinct queries (M_i, M_j) such that $F(M_i) = F(M_j)$ is greater than $\bar{q}_c/4$,
  - $\mathsf{Bad}_{5b} \Leftrightarrow$ the number of distinct queries (M_i, M_j) such that $G(M_i) = G(M_j)$ is greater than $\bar{q}_c/4$.

We distinguish two types of trails of length 4; a trail of type M consists of two F-collisions and one G-collision, while a trail of type W consists of two G-collisions and one F-collision. Then we observe that

1. $\mathcal{G}_\tau^=$ contains no cycle of length 2 without Bad_1;
2. $\mathcal{G}_\tau^=$ contains no trail of type M without Bad_2;
3. $\mathcal{G}_\tau^=$ contains no trail of type W whose label is $\mathbf{0}$ without Bad_3;
4. $\mathcal{G}_\tau^=$ contains no trail of length 2 whose label is $\mathbf{0}$ without Bad_4;

5. the number of edges in $\mathcal{C}$ is not greater than $\bar{q}_c$ without Bad_5.

Without Bad_2, $\mathcal{G}_\tau^=$ contains neither a cycle of length 4 nor a trail of length 5. We also note that $\lambda(\mathcal{L}) \neq \mathbf{0}$ for any trail $\mathcal{L}$ of even length without Bad_2, Bad_3 and Bad_4. Therefore, we have

$$\Pr[\mathsf{T}_{\mathrm{id}} \in \mathit{\Gamma}_{\mathsf{bad}}] \leq \Pr[\mathsf{Bad}_1 \vee \mathsf{Bad}_2 \vee \mathsf{Bad}_3 \vee \mathsf{Bad}_4 \vee \mathsf{Bad}_5].$$

Auxiliary Events. For each $i = 1, \ldots, q$, the i-th message is denoted $M_i = M_i[1] \| \cdots \| M_i[m_i]$, where m_i is the length of M_i in blocks. For distinct $i, j \in [q]$, let

$$\begin{aligned}\mathsf{NEQ}_{i,j} \stackrel{\text{def}}{=} &\{\alpha \in [\min\{m_i, m_j\}] : M_i[\alpha] \neq M_j[\alpha]\} \\ &\sqcup \{\alpha : \min\{m_i, m_j\} < \alpha \leq \max\{m_i, m_j\}\}.\end{aligned}$$

Since $M_i[\alpha] = M_j[\alpha]$ for any index $\alpha \notin \mathsf{NEQ}_{i,j}$, we can simply ignore such an index when we consider F- and G-collisions. We also note that $\mathsf{NEQ}_{i,j} \neq \emptyset$ if M_i and M_j are distinct.

Once $\mathit{\Delta}_0 = \pi(0)$ and $\mathit{\Delta}_1 = \pi(1)$ are fixed, we obtain $X_i = X_i[1] \| \cdots \| X_i[m_i]$, where $X_i[\alpha] = M_i[\alpha] \oplus 2^\alpha \cdot \mathit{\Delta}_0 \oplus 2^{2\alpha} \cdot \mathit{\Delta}_1$. Let

$$\mathcal{I}_{\mathsf{col}} \stackrel{\text{def}}{=} \{(i,j) \in [q]^{*2} : X_i[\alpha] = X_j[\beta] \text{ for some } \alpha,\ \beta \text{ such that } \alpha \neq \beta\},$$

$$\mathcal{I}'_{\mathsf{col}} \stackrel{\text{def}}{=} \{(i,j) \in [q]^{*2} : \min\{\mathsf{NEQ}_{i,j}\} \leq m_i \text{ and } X_i[\min\{\mathsf{NEQ}_{i,j}\}] = X_j[\beta] \text{ for some } \beta\}.$$

In order to analyze the probability of the bad events, we need to introduce certain auxiliary events as follows.

- $\mathsf{Aux}_1 \Leftrightarrow$ either $\pi(0) = 0$ or $\pi(1) = 0$;
- $\mathsf{Aux}_2 \Leftrightarrow X_i[\alpha] = X_i[\beta]$ for some $i \in [q]$ and two distinct indices α and β;
- $\mathsf{Aux}_3 \Leftrightarrow X_i[\alpha] \in \{0, 1, \pi^{-1}(0)\}$ for some $i \in [q]$ and $\alpha \in [m_i]$;
- $\mathsf{Aux}_4 \Leftrightarrow |\mathcal{I}_{\mathsf{col}}| > \hat{q}_c$;
- $\mathsf{Aux}_5 \Leftrightarrow |\mathcal{I}'_{\mathsf{col}}| > \bar{q}_c$.

Note that $\bar{q}_c$ has been introduced in Section 3, while $\hat{q}_c$ is a new one. Let $\mathsf{Aux} = \mathsf{Aux}_1 \vee \mathsf{Aux}_2 \vee \mathsf{Aux}_3 \vee \mathsf{Aux}_4 \vee \mathsf{Aux}_5$. It is not hard to see that if $\ell \leq N$, then

$$\Pr[\mathsf{Aux}_1 \vee \mathsf{Aux}_3] \leq \frac{3\ell q}{N-2} + \frac{2}{N}, \qquad \Pr[\mathsf{Aux}_2] \leq \frac{\ell^2 q}{2N},$$
$$\Pr[\mathsf{Aux}_4] \leq \frac{\ell^2 q^2}{\hat{q}_c N}, \qquad \Pr[\mathsf{Aux}_5] \leq \frac{\ell q^2}{\bar{q}_c N}$$

over the random choice of $\pi(0)$, $\pi(1)$, $\pi^{-1}(0)$. Simplifying the bounds, we have

$$\Pr[\mathsf{Aux}] \leq \frac{(\ell^2 + 8\ell)q}{2N} + \frac{\ell^2 q^2}{\hat{q}_c N} + \frac{\ell q^2}{\bar{q}_c N}. \tag{9}$$

Almost Universality. The almost universality of each half of PHash will be used to upper bound the probability of Bad_4 and Bad_5.

Lemma 2. *Let* $\mathsf{PHash}(M) = (F(M), G(M))$ *for any message* M*. If* $\ell \leq N/4$*, then* F *and* G *are* δ*-almost universal, where*

$$\delta = \frac{8\ell}{N}.$$

We refer to [19] for the proof of Lemma 2.

Classifying X-Variables. In order to upper bound the probability of Bad_1, Bad_2, Bad_3, we need to classify X-variables for each pair of messages, assuming that Aux has not occurred; let

$$\mathcal{X}_{i,j} = \mathcal{X}_{\bar{i},j} \sqcup \mathcal{X}_{i,\bar{j}} \sqcup \mathcal{X}_{\bar{i},\bar{j}}$$

where

$$\begin{aligned}
\mathcal{X}_{\bar{i},j} &\stackrel{\text{def}}{=} \{X_i[\alpha] : \alpha \in \mathsf{NEQ}_{i,j}\} \setminus \{X_j[\alpha] : \alpha \in \mathsf{NEQ}_{i,j}\},\\
\mathcal{X}_{i,\bar{j}} &\stackrel{\text{def}}{=} \{X_j[\alpha] : \alpha \in \mathsf{NEQ}_{i,j}\} \setminus \{X_i[\alpha] : \alpha \in \mathsf{NEQ}_{i,j}\},\\
\mathcal{X}_{\bar{i},\bar{j}} &\stackrel{\text{def}}{=} \{X_i[\alpha] : \alpha \in \mathsf{NEQ}_{i,j}\} \cap \{X_j[\alpha] : \alpha \in \mathsf{NEQ}_{i,j}\}.
\end{aligned}$$

We make the following observations.

1. If $X \in \mathcal{X}_{\bar{i},\bar{j}}$, then we have $X = X_i[\alpha] = X_j[\beta]$ for distinct indices α and β.
2. If $\mathcal{X}_{\bar{i},j} \cup \mathcal{X}_{i,\bar{j}} = \emptyset$, then $F(M_i) = F(M_j)$ (regardless of π); the probability that $\mathcal{X}_{\bar{i},j} \cup \mathcal{X}_{i,\bar{j}} = \emptyset$ is upper bounded by $\frac{\ell}{N-1}$ over the random choice of Δ_0 and Δ_1.
3. If $\mathcal{X}_{\bar{i},j} \cup \mathcal{X}_{i,\bar{j}}$ contains either one or two elements, then it is not possible that $F(M_i) = F(M_j)$.
4. The probability that $\mathcal{X}_{\bar{i},\bar{j}} \neq \emptyset$ is upper bounded by $\frac{\ell^2}{N-1}$ over the random choice of Δ_0 and Δ_1.

By relabeling, let

$$\begin{aligned}
\mathcal{X}_{i,j} &= \{X[1], \ldots, X[t]\},\\
\mathcal{Y}_{i,j} &= \{Y[1], \ldots, Y[t]\},
\end{aligned}$$

where $t = |\mathcal{X}_{i,j}|$ and $Y[\alpha] = \pi(X[\alpha])$ for $\alpha = 1, \ldots, t$. We also partition the set of indices $\{1, \ldots, t\}$ into three subsets; $\{1, \ldots, t\} = I_{\bar{i},j} \sqcup I_{i,\bar{j}} \sqcup I_{\bar{i},\bar{j}}$, where

$$\begin{aligned}
\alpha \in I_{\bar{i},j} &\Leftrightarrow X[\alpha] \in \mathcal{X}_{\bar{i},j},\\
\alpha \in I_{i,\bar{j}} &\Leftrightarrow X[\alpha] \in \mathcal{X}_{i,\bar{j}},\\
\alpha \in I_{\bar{i},\bar{j}} &\Leftrightarrow X[\alpha] \in \mathcal{X}_{\bar{i},\bar{j}}.
\end{aligned}$$

Then we can represent F- and G-collisions by equations in $Y[\alpha]$ as follows.

$$F(M_i) = F(M_j) \Leftrightarrow A_1 \cdot Y[1] \oplus \cdots \oplus A_t \cdot Y[t] = 0, \tag{10}$$

$$G(M_i) = G(M_j) \Leftrightarrow B_1 \cdot Y[1] \oplus \cdots \oplus B_t \cdot Y[t] = 0, \tag{11}$$

where

1. $A_\alpha = 1$ if $\alpha \in I_{\bar{i},j} \cup I_{i,\bar{j}}$, and $A_\alpha = 0$ if $\alpha \in I_{\bar{i},\bar{j}}$;
2. $B_\alpha = 2^\beta$ for some β if $\alpha \in I_{\bar{i},j} \cup I_{i,\bar{j}}$, and $B_\alpha = 2^\beta \oplus 2^\gamma$ for distinct β and γ if $\alpha \in I_{\bar{i},\bar{j}}$.

Each unknown $Y[\alpha]$ can be seen as a random variable whose value is taken from a set of size $N-3$, namely $\{0,1\}^n \setminus \{0, \pi(0), \pi(1)\}$.

Upper Bounding the Probability of Bad Events. We are now ready to upper bound the probability of each bad event above.

Lemma 3. *Assume that $\ell \le \frac{N}{8}$. Then, in the ideal world, one has*

$$\Pr[\mathsf{Bad}_1 \wedge \neg\mathsf{Aux}] \le \frac{4\ell q^2}{N^2}.$$

Proof. We fix distinct $i, j \in [q]$, and distinguish the following two cases.

Case 1: $\mathcal{X}_{\bar{i},j} \cup \mathcal{X}_{i,\bar{j}} = \emptyset$. This case happens with probability at most $\frac{\ell}{N-1}$ over the random choice of Δ_0 and Δ_1. Since all the coefficients B_α in (11) are nonzero, the probability that $G(M_i) = G(M_j)$ is upper bounded by $(N-3)_{t-1}/(N-3)_t$, which is not greater than $\frac{1}{N-2\ell-2}$ since $t \le 2\ell$.

Case 2: $\mathcal{X}_{\bar{i},j} \cup \mathcal{X}_{i,\bar{j}} \neq \emptyset$. It should be the case that $|\mathcal{X}_{\bar{i},j} \cup \mathcal{X}_{i,\bar{j}}| \ge 2$ since otherwise we have $F(M_i) \neq F(M_j)$. Consider equations (10) and (11) (with the same pair of i and j). There are at least two indices α, $\alpha' \in I_{\bar{i},j} \cup I_{i,\bar{j}}$, where $A_\alpha = A_{\alpha'} = 1$, $B_\alpha = 2^\beta$ and $B_{\alpha'} = 2^\gamma$ for distinct β and γ. So the system of equations has rank 2, and hence the equations are satisfied with probability at most $(N-3)_{t-2}/(N-3)_t$, which is not greater than $\frac{1}{(N-2\ell-1)(N-2\ell-2)}$.

Overall, we have $\Pr[\mathsf{Bad}_1 \wedge \neg\mathsf{Aux}] \le \frac{4\ell q^2}{N^2}$ since $\ell \le \frac{N}{8}$. □

Lemma 4. *Assume that $\ell \le \frac{N}{16}$. Then, in the ideal world, one has*

$$\Pr[\mathsf{Bad}_2 \wedge \neg\mathsf{Aux}] \le \frac{2\bar{q_c}^2}{N} + \frac{4\hat{q_c}}{N} + \frac{2}{N} + \frac{2\sqrt{2}q^2}{N^{\frac{3}{2}}} + \frac{8\hat{q_c}q^2}{N^2} + \frac{96q^2}{N^2} + \frac{8q^4}{N^3}.$$

Proof. We partition the set $[q]^{*4}$ of quadruples into five subsets; $[q]^{*4} = \mathcal{J}_1 \sqcup \mathcal{J}_2 \sqcup \mathcal{J}_3 \sqcup \mathcal{J}_4 \sqcup \mathcal{J}_5$, where

$$\mathcal{J}_1 \stackrel{\text{def}}{=} \left\{(i_1,i_2,i_3,i_4) \in [q]^{*4} : (i_2,i_3) \in \mathcal{I}_{\mathsf{col}}\right\},$$
$$\mathcal{J}_2 \stackrel{\text{def}}{=} \left\{(i_1,i_2,i_3,i_4) \in [q]^{*4} : (i_2,i_3) \notin \mathcal{I}_{\mathsf{col}} \wedge (i_1,i_2) \in \mathcal{I}_{\mathsf{col}} \wedge (i_3,i_4) \in \mathcal{I}_{\mathsf{col}}\right\},$$
$$\mathcal{J}_3 \stackrel{\text{def}}{=} \left\{(i_1,i_2,i_3,i_4) \in [q]^{*4} : (i_2,i_3) \notin \mathcal{I}_{\mathsf{col}} \wedge (i_1,i_2) \notin \mathcal{I}_{\mathsf{col}} \wedge (i_3,i_4) \in \mathcal{I}_{\mathsf{col}}\right\},$$
$$\mathcal{J}_4 \stackrel{\text{def}}{=} \left\{(i_1,i_2,i_3,i_4) \in [q]^{*4} : (i_2,i_3) \notin \mathcal{I}_{\mathsf{col}} \wedge (i_1,i_2) \in \mathcal{I}_{\mathsf{col}} \wedge (i_3,i_4) \notin \mathcal{I}_{\mathsf{col}}\right\},$$
$$\mathcal{J}_5 \stackrel{\text{def}}{=} \left\{(i_1,i_2,i_3,i_4) \in [q]^{*4} : (i_2,i_3) \notin \mathcal{I}_{\mathsf{col}} \wedge (i_1,i_2) \notin \mathcal{I}_{\mathsf{col}} \wedge (i_3,i_4) \notin \mathcal{I}_{\mathsf{col}}\right\}.$$

For $(i_1,i_2,i_3,i_4) \in [q]^{*4}$, let

$$\mathsf{Bad}_2^{i_1,i_2,i_3,i_4} \Leftrightarrow F(M_{i_1}) = F(M_{i_2}) \wedge G(M_{i_2}) = G(M_{i_3}) \wedge F(M_{i_3}) = F(M_{i_4}).$$

Then we have

$$\mathsf{Bad}_2 \Leftrightarrow \bigvee_{(i_1,i_2,i_3,i_4)\in[q]^{*4}} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4},$$

and hence,

$$\Pr\left[\mathsf{Bad}_2 \wedge \neg\mathsf{Aux}\right] \le \mathsf{p}_1 + \mathsf{p}_2 + \mathsf{p}_3 + \mathsf{p}_4 + \mathsf{p}_5,$$

where

$$\mathsf{p}_j \stackrel{\text{def}}{=} \Pr\left[\left(\bigvee_{(i_1,i_2,i_3,i_4)\in\mathcal{J}_j} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right) \wedge \neg\mathsf{Aux}\right]$$

for $j = 1, 2, 3, 4, 5$.

For a fixed quadruple (i_1, i_2, i_3, i_4), we can represent $\mathsf{Bad}_2^{i_1,i_2,i_3,i_4}$ by a system of three linear equations;

$$\begin{aligned}
F(M_{i_1}) = F(M_{i_2}) &\Leftrightarrow A_{1,1}\cdot Y[1] \oplus \cdots \oplus A_{1,t}\cdot Y[t] = 0,\\
G(M_{i_2}) = G(M_{i_3}) &\Leftrightarrow A_{2,1}\cdot Y[1] \oplus \cdots \oplus A_{2,t}\cdot Y[t] = 0,\\
F(M_{i_3}) = F(M_{i_4}) &\Leftrightarrow A_{3,1}\cdot Y[1] \oplus \cdots \oplus A_{3,t}\cdot Y[t] = 0
\end{aligned}$$

for some $A_{j,\alpha}$, where each column corresponds to a variable in

$$\mathcal{X}_{\bar{i_1},i_2} \cup \mathcal{X}_{i_1,\bar{i_2}} \cup \mathcal{X}_{i_2,i_3} \cup \mathcal{X}_{\bar{i_3},i_4} \cup \mathcal{X}_{i_3,\bar{i_4}},$$

so the number of columns, denoted t, is the size of this set. This system of equations can also be regarded as a $3 \times t$ matrix $(A_{j,\alpha})$. This matrix will sometimes be denoted $A^{(i_1,i_2,i_3,i_4)}$ to specify the corresponding quadruple. For $j = 1, 2, 3$, the j-th row of $(A_{j,\alpha})$ is denoted $A_j^{(i_1,i_2,i_3,i_4)}$, or simply A_j. We observe that the second row A_2 is always nonzero, namely, the G-collision is nontrivial.

Upper Bounding p_1. We have

$$\left(\bigvee_{(i_1,i_2,i_3,i_4)\in\mathcal{J}_1} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right) \wedge \neg\mathsf{Aux} \Rightarrow \left(\bigvee_{(i_2,i_3)\in\mathcal{I}_{\mathsf{col}}} G(M_{i_2}) = G(M_{i_3})\right) \wedge \neg\mathsf{Aux}.$$

Since $|\mathcal{I}_{\mathsf{col}}| \le \hat{q}_c$ and the G-collision is nontrivial, the probability of the event on the right-hand side is upper bounded by $\hat{q}_c/(N - 2\ell - 2)$. So we have

$$\mathsf{p}_1 \le \frac{2\hat{q}_c}{N}. \tag{12}$$

Upper Bounding p_2. We have

$$\left(\bigvee_{(i_1,i_2,i_3,i_4)\in\mathcal{J}_2} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right) \wedge \neg\mathsf{Aux} \Rightarrow \left(\bigvee_{(i_1,i_2)\in\mathcal{I}_{\mathsf{col}}\setminus\mathcal{I}'_{\mathsf{col}}} F(M_{i_1}) = F(M_{i_2})\right)$$
$$\wedge \left(\bigvee_{\substack{(i_1,i_2)\in\mathcal{I}'_{\mathsf{col}}\\(i_3,i_4)\in\mathcal{I}'_{\mathsf{col}}\\(i_2,i_3)\notin\mathcal{I}_{\mathsf{col}}}} G(M_{i_2}) = G(M_{i_3})\right) \wedge \neg\mathsf{Aux}$$

We see that

1. for any pair of messages in $\mathcal{I}_{\mathsf{col}} \setminus \mathcal{I}'_{\mathsf{col}}$, their F-collision is nontrivial;
2. for any pair of messages in $[q]^{*2} \setminus \mathcal{I}_{\mathsf{col}}$, their G-collision is nontrivial;
3. $|\mathcal{I}_{\mathsf{col}} \setminus \mathcal{I}'_{\mathsf{col}}| \leq \hat{q}_c$ and $|\mathcal{I}'_{\mathsf{col}}| \leq \bar{q}_c$.

Therefore we have

$$\mathsf{p}_2 \leq \frac{\hat{q}_c}{N-2\ell-2} + \frac{\bar{q}_c{}^2}{N-2\ell-2} \leq \frac{2\hat{q}_c}{N} + \frac{2\bar{q}_c{}^2}{N}. \tag{13}$$

Upper Bounding p_3. Fix a quadruple $(i_1, i_2, i_3, i_4) \in \mathcal{J}_3$, and consider the corresponding matrix $A^{(i_1,i_2,i_3,i_4)} = (A_{j,\alpha})$. A_1 is a zero-one matrix, but nonzero since $(i_1, i_2) \notin \mathcal{I}_{\mathsf{col}}$, while A_2 contains at least two entries, say 2^β and 2^γ for distinct β and γ. This implies that A_2 cannot be a multiple of A_1, and hence $(A_{j,\alpha})$ has rank at least two. Therefore the probability that random variables $Y[1], \ldots, Y[t]$ satisfy the system of equations is upper bounded by $(N-3)_{t-2}/(N-3)_t$, which is $1/(N-t-1)(N-t-2)$. Since the number of quadruples $(i_1, i_2, i_3, i_4) \in [q]^{*4}$ such that $(i_2, i_3) \notin \mathcal{I}_{\mathsf{col}}$ is at most $\hat{q}_c q^2$ and since $t \leq 4\ell$, we have

$$\mathsf{p}_3 \leq \frac{\hat{q}_c q^2}{(N-4\ell-1)(N-4\ell-2)} \leq \frac{4\hat{q}_c q^2}{N^2}. \tag{14}$$

Upper Bounding p_4. In a similar manner to the analysis of p_3, we obtain

$$\mathsf{p}_4 \leq \frac{\hat{q}_c q^2}{(N-4\ell-1)(N-4\ell-2)} \leq \frac{4\hat{q}_c q^2}{N^2}. \tag{15}$$

Upper Bounding p_5. Fix a quadruple $(i_1, i_2, i_3, i_4) \in \mathcal{J}_5$, and consider the corresponding matrix $A^{(i_1,i_2,i_3,i_4)} = (A_{j,\alpha})$. We can assume that A_1 and A_3 contain at least three 1's, since otherwise we will not have two F-collisions for A_1 and A_3. Every entry of A_2 should be given as 2^α for some α (since $(i_2, i_3) \notin \mathcal{I}_{\mathsf{col}}$), and for each α, 2^α appears at most twice in the row. Furthermore, A_2 should contain at least two distinct entries, since otherwise we will not have the G-collision (with distinct nonzero Y-variables). So A_2 cannot be a multiple of A_1, and hence the rank of $(A_{j,\alpha})$ is at least two. In this case, we have two possibilities; one is that $A_1 = A_3$, and the other is that $A_2 = CA_1 \oplus DA_3$ for some nonzero constants C and D.

All in all, $\mathcal{J}_5$ can be represented by a union of three subsets; $\mathcal{J}_5 = \mathcal{J}_{5,1} \cup \mathcal{J}_{5,2} \cup \mathcal{J}_{5,3}$, where

$$\mathcal{J}_{5,1} \stackrel{\text{def}}{=} \left\{(i_1, i_2, i_3, i_4) \in \mathcal{J}_5 : A^{(i_1,i_2,i_3,i_4)} \text{ has rank } 3\right\},$$
$$\mathcal{J}_{5,2} \stackrel{\text{def}}{=} \left\{(i_1, i_2, i_3, i_4) \in \mathcal{J}_5 : A_1^{(i_j)} = A_3^{(i_j)}\right\},$$
$$\mathcal{J}_{5,3} \stackrel{\text{def}}{=} \left\{(i_1, i_2, i_3, i_4) \in \mathcal{J}_5 : A_2^{(i_j)} = CA_1^{(i_j)} \oplus DA_3^{(i_j)} \text{ for nonzero } C \text{ and } D\right\}.$$

For $(i_1, i_2, i_3, i_4) \in \mathcal{J}_{5,1}$, it is not hard to see that the probability of Y-variables satisfying the corresponding system of equations is upper bounded by $(N-3)_{t-3}/(N-3)_t$, which is $1/(N-t)(N-t-1)(N-t-2)$. Since $t \leq 4\ell$, we have

$$\Pr\left[\left(\bigvee_{(i_1,i_2,i_3,i_4)\in\mathcal{J}_{5,1}} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right) \wedge \neg\mathsf{Aux}\right] \leq \frac{q^4}{(N-4\ell)(N-4\ell-1)(N-4\ell-2)} \leq \frac{8q^4}{N^3}. \quad (16)$$

In order to upper bound the probability of $\mathsf{Bad}_2^{i_1,i_2,i_3,i_4}$ for $(i_1, i_2, i_3, i_4) \in \mathcal{J}_{5,2}$, we need to define an equivalence relation, denoted $\sim$, on $[q]^{*2} \setminus \mathcal{I}_{\mathsf{col}}$, where

$$(i_1, i_2) \sim (i_3, i_4) \Leftrightarrow \mathcal{X}_{\bar{i_1},i_2} \sqcup \mathcal{X}_{i_1,\bar{i_2}} = \mathcal{X}_{\bar{i_3},i_4} \sqcup \mathcal{X}_{i_3,\bar{i_4}}.$$

The relation $(i_1, i_2) \sim (i_3, i_4)$ implies that $A_1 = A_3$ for $A^{(i_1,i_2,i_3,i_4)}$. In other words, $F(M_{i_1}) = F(M_{i_2}) \Leftrightarrow F(M_{i_3}) = F(M_{i_4})$, namely, the two F-collisions are dependent on each other. We will assume that this relation partitions $[q]^{*2} \setminus \mathcal{I}_{\mathsf{col}}$ into r subsets, denoted $\mathcal{I}_1, \ldots, \mathcal{I}_r$, respectively. So we have

$$[q]^{*2} \setminus \mathcal{I}_{\mathsf{col}} = \mathcal{I}_1 \sqcup \cdots \sqcup \mathcal{I}_r.$$

For $j = 1, \ldots, r$, let

$$\mathsf{E}_j \Leftrightarrow F(M_{i_1}) = F(M_{i_2}) \text{ for every } (i_1, i_2) \in \mathcal{I}_j.$$

Then we have

$$\Pr\left[\mathsf{E}_j \wedge \neg\mathsf{Aux}\right] \leq \frac{1}{N-2\ell-2}.$$

Given $\neg\mathsf{Aux}$, we have

$$\left(\bigvee_{(i_1,i_2,i_3,i_4)\in\mathcal{J}_{5,2}} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right) \Rightarrow \left(\bigvee_{j\in[r]} \bigvee_{(i_1,i_2),(i_3,i_4)\in\mathcal{I}_j} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right).$$

For each $j = 1, \ldots, r$, we have

$$\begin{aligned}
&\Pr\left[\left(\bigvee_{(i_1,i_2),(i_3,i_4)\in\mathcal{I}_j} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right) \wedge \neg\mathsf{Aux}\right] \\
&\leq \Pr\left[\mathsf{E}_j \wedge \neg\mathsf{Aux}\right] \cdot \Pr\left[\bigvee_{(i_1,i_2),(i_3,i_4)\in\mathcal{I}_j} G(M_{i_2}) = G(M_{i_3}) \,\middle|\, \mathsf{E}_j \wedge \neg\mathsf{Aux}\right] \\
&\leq \frac{1}{N-2\ell-2} \cdot \min\left(\frac{|\mathcal{I}_j|^2}{N-3\ell-2}, 1\right)
\end{aligned}$$

since the first and the second rows of $A^{(i_1,i_2,i_3,i_4)}$ are always linearly independent. Overall, we have

$$\Pr\left[\left(\bigvee_{(i_1,i_2,i_3,i_4)\in\mathcal{J}_{5,2}} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right)\wedge\neg\mathsf{Aux}\right] \le \sum_{j=1}^{r}\frac{2}{N}\cdot\min\left(\frac{2\left|\mathcal{I}_j\right|^2}{N},1\right) \quad (17)$$

where we use $\ell \le N/16$. Subject to the condition $\sum_{j=1}^{r}|\mathcal{I}_j| = q^2$ (and with no restriction on r), $\sum_{j=1}^{r}\min\left(\frac{2|\mathcal{I}_j|^2}{N},1\right)$ is maximized when $r = \left\lfloor q^2/\left(\frac{N}{2}\right)^{\frac{1}{2}}\right\rfloor + 1$, $|\mathcal{I}_j| = \left(\frac{N}{2}\right)^{\frac{1}{2}}$ for $j = 1,\ldots,r-1$ and $|\mathcal{I}_r| = q^2 - (r-1)\left(\frac{N}{2}\right)^{\frac{1}{2}}$, in which case we have

$$\sum_{j=1}^{r}\frac{2}{N}\cdot\min\left(\frac{2\left|\mathcal{I}_j\right|^2}{N},1\right) \le \frac{2\sqrt{2}q^2}{N^{\frac{3}{2}}} + \frac{2}{N}. \quad (18)$$

Finally, we focus on $A^{(i_1,i_2,i_3,i_4)}$ for $(i_1,i_2,i_3,i_4)\in\mathcal{J}_{5,3}$. We note that A_2 is represented by a linear combination of A_1 and A_3, where we can assume that

1. A_2 does not contain the same entry more than twice;
2. A_2 contains at least two different nonzero entries;
3. each of A_1 and A_3 contains at least three 1's.

Therefore the supports of A_1 and A_3 cannot intersect at more than two positions, nor be disjoint each other. So we should be able to find a 3×3 submatrix

$$\begin{bmatrix} 1 & 1 & 0 \\ C & C\oplus D & D \\ 0 & 1 & 1 \end{bmatrix}$$

where $C = 2^\alpha$ and $D = 2^\beta$ for distinct α and β. Furthermore, it should be the case that $2^\alpha \oplus 2^\beta = 2^\gamma$ for some γ since $(i_2,i_3)\notin\mathcal{I}_{\mathsf{col}}$. Since a linear combination of A_1 and A_3 generates at most three different nonzero values in A_2, we conclude that $\mathsf{NEQ}_{i_2,i_3} = \{\alpha,\beta,\gamma\}$.

Suppose that we begin with two messages M_{i_2} and M_{i_3} such that $|\mathsf{NEQ}_{i_2,i_3}| = 3$, and try to find M_{i_1} and M_{i_4} such that $(i_1,i_2,i_3,i_4)\in\mathcal{J}_{5,3}$. Let $\mathsf{NEQ}_{i_2,i_3} = \{\alpha,\beta,\gamma\}$, where $2^\alpha\oplus 2^\beta\oplus 2^\gamma = 0$ and $\alpha<\beta<\gamma$. Then A_2 is uniquely determined by M_{i_2} and M_{i_3}, and its nonzero elements are 2^α, 2^β, 2^γ, each of which appears once or twice in the row. Once we choose a pair of distinct coefficients $(C,D)\in\{2^\alpha,2^\beta,2^\gamma\}^{*2}$, we can fix A_1 and A_3 such that $CA_1\oplus DA_3 = A_2$. For example, if every nonzero element appears exactly twice in A_2, and if $C = 2^\alpha$ and $D = 2^\beta$, then A will contain a 3×6 submatrix

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 \\ 2^\alpha & 2^\beta & 2^\gamma & 2^\alpha & 2^\beta & 2^\gamma \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

with all the other entries being zero. Since we have at most two possibilities for M_{i_1} (resp. M_{i_4}) yielding A_1 (resp. A_3), the number of possible ways of choosing

M_{i_1} and M_{i_4} is at most 24 (given M_{i_2} and M_{i_3}), and for each of such quadruples, the probability that the Y-variables satisfy the corresponding system of equations is upper bounded by $1/(N-4\ell-1)(N-4\ell-2)$. Therefore we have

$$\Pr\left[\left(\bigvee_{(i_1,i_2,i_3,i_4)\in\mathcal{J}_{5,3}} \mathsf{Bad}_2^{i_1,i_2,i_3,i_4}\right)\wedge\neg\mathsf{Aux}\right] \leq \frac{24q^2}{(N-4\ell-1)(N-4\ell-2)} \leq \frac{96q^2}{N^2}. \quad (19)$$

By (16), (17), (18), (19), we have

$$\mathsf{p}_5 \leq \frac{2}{N} + \frac{2\sqrt{2}q^2}{N^{\frac{3}{2}}} + \frac{96q^2}{N^2} + \frac{8q^4}{N^3}. \quad (20)$$

The proof is now complete by (12), (13), (14), (15), (20). □

Lemma 5. *Assume that $\ell \leq \frac{N}{8}$. Then, in the ideal world, one has*

$$\Pr[\mathsf{Bad}_3 \wedge \neg\mathsf{Aux}] \leq \frac{6\ell^2q^4}{N^3}.$$

Proof. Fix a quadruple of distinct queries. For simplicity of notation and without loss of generality, we will consider (M_1, M_2, M_3, M_4). In the ideal world, the probability that $T_1 \oplus T_2 \oplus T_3 \oplus T_4 = \mathbf{0}$ is $\frac{1}{N}$.

Next, we will upper bound the probability that $F(M_1) = F(M_2)$ and $G(M_2) = G(M_3)$, focusing on the first three messages. We consider the following three cases.

Case 1: $\mathcal{X}_{\bar{1},2} \cup \mathcal{X}_{1,\bar{2}} = \emptyset$. The analysis is similar to Case 1 in Lemma 3; the probability that $F(M_1) = F(M_2)$ and $G(M_2) = G(M_3)$ in this case is upper bounded by $\frac{\ell}{(N-1)(N-2\ell-2)}$.

Case 2: $\mathcal{X}_{\bar{1},2} \cup \mathcal{X}_{1,\bar{2}} \neq \emptyset$ and $\mathcal{X}_{\bar{2},\bar{3}} \neq \emptyset$. The probability that $\mathcal{X}_{\bar{2},\bar{3}} \neq \emptyset$ (over the random choice of Δ_0 and Δ_1) is upper bounded by $\frac{\ell^2}{N-1}$. Once Δ_0 and Δ_1 are fixed, the probability that $F(M_1) = F(M_2)$ (over the random choice of π) is upper bounded by $\frac{1}{N-2\ell-2}$.

Case 3: $\mathcal{X}_{\bar{1},2} \cup \mathcal{X}_{1,\bar{2}} \neq \emptyset$ and $\mathcal{X}_{\bar{2},\bar{3}} = \emptyset$. It should be the case that $|\mathcal{X}_{\bar{1},2} \cup \mathcal{X}_{1,\bar{2}}| \geq 2$. The F- and G-collisions can be represented by a system of equations

$$A_{1,1} \cdot Y[1] \oplus \cdots \oplus A_{1,t} \cdot Y[t] = 0,$$
$$A_{2,1} \cdot Y[1] \oplus \cdots \oplus A_{2,t} \cdot Y[t] = 0,$$

for some $A_{j,\alpha}$, where $t = |\mathcal{X}_{\bar{1},2} \cup \mathcal{X}_{1,\bar{2}} \cup \mathcal{X}_{2,3}|$. We can also partition the set of indices $\{1,\ldots,t\}$ into two subsets; $\{1,\ldots,t\} = I_1 \sqcup I_2$, where

$$\alpha \in I_1 \Leftrightarrow X[\alpha] \in \mathcal{X}_{\bar{1},2} \sqcup \mathcal{X}_{1,\bar{2}},$$
$$\alpha \in I_2 \Leftrightarrow X[\alpha] \in \mathcal{X}_{2,3} \setminus (\mathcal{X}_{\bar{1},2} \cup \mathcal{X}_{1,\bar{2}}).$$

We note that $A_{1,\alpha} = 1$ for every $\alpha \in I_1$ and $A_{1,\alpha} = 0$ for every $\alpha \in I_2$. Furthermore, for every $\alpha \in I_2$, $A_{2,\alpha}$ is nonzero. So if I_2 is nonempty, then $(A_{i,\alpha})$ contains a 2×2 submatrix

$$\begin{bmatrix} 1 & 0 \\ * & 2^\beta \end{bmatrix}$$

for some β, and hence the system of equations has rank 2.

If I_2 is empty, then $\mathcal{X}_{\bar{2},3} \cup \mathcal{X}_{2,\bar{3}} \subset \mathcal{X}_{\bar{1},2} \sqcup \mathcal{X}_{1,\bar{2}}$. We also have $|\mathcal{X}_{\bar{2},3} \cup \mathcal{X}_{2,\bar{3}}| \geq 2$ since otherwise $G(M_2) \neq G(M_3)$. So we have two indices α, $\alpha' \in I_1$ such that $X[\alpha]$, $X[\alpha'] \in \mathcal{X}_{\bar{2},3} \cup \mathcal{X}_{2,\bar{3}}$. Since $A_{2,\alpha} = 2^\beta$ and $A_{2,\alpha'} = 2^\gamma$ for distinct β and γ, $(A_{i,\alpha})$ contains a 2×2 submatrix

$$\begin{bmatrix} 1 & 1 \\ 2^\beta & 2^\gamma \end{bmatrix}$$

for distinct β and γ, and hence the system of equations has rank 2. So in any case, the system of equations are satisfied with probability at most $\frac{1}{(N-2\ell-1)(N-2\ell-2)}$. Overall, we have $\Pr[\mathsf{Bad}_3 \wedge \neg\mathsf{Aux}] \leq \frac{6\ell^2 q^4}{N^3}$ since $\ell \leq \frac{N}{8}$. □

The following two lemmas are easy to prove using the Markov inequality and the almost universality of F and G.

Lemma 6. *In the ideal world, one has*

$$\Pr[\mathsf{Bad}_4] \leq \frac{16\ell^2 q^2}{N^2}.$$

Lemma 7. *In the ideal world, one has*

$$\Pr[\mathsf{Bad}_5] \leq \frac{64\ell q^2}{\bar{q}_c N}.$$

By Lemma 3, 4, 5, 6, 7, and (9), we can upper bound the probability of Bad, and then combining it with (7) (setting $\hat{q}_c = \ell N^{\frac{1}{2}}/2\sqrt{2}$ and $\bar{q}_c = 2\ell^{\frac{1}{3}} q^{\frac{2}{3}}$), we obtain the following theorem.

Theorem 5. *Assume that $\ell \leq N/16$. When* PMAC-Plus *is based on a block cipher E, one has*

$$\begin{aligned}\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{PMAC\text{-}Plus}}(q,t,\ell) \leq{}& \frac{\sqrt{2}\ell}{N^{\frac{1}{2}}} + \frac{45\ell^{\frac{2}{3}} q^{\frac{4}{3}}}{N} + \frac{(\ell^2+8\ell)q}{2N} + \frac{2}{N} + \frac{(4\sqrt{2}\ell + 2\sqrt{2})q^2}{N^{\frac{3}{2}}} \\ &+ \frac{3\ell^{\frac{1}{3}} q^{\frac{8}{3}}}{N^2} + \frac{9\ell^{\frac{2}{3}} q^{\frac{7}{3}}}{2N^2} + \frac{(16\ell^2+4\ell+97)q^2}{N^2} + \frac{(18\ell^2+32)q^4}{3N^3} \\ &+ 3\mathbf{Adv}^{\mathsf{prp}}_E(\ell q, t+t'),\end{aligned}$$

where t' is the time complexity necessary to compute E for ℓq times.

Note that all the constant coefficients are loosely estimated in our bounds; most large coefficients appear since we replace $N - c\ell$ by $N/2$ for any small integer c.

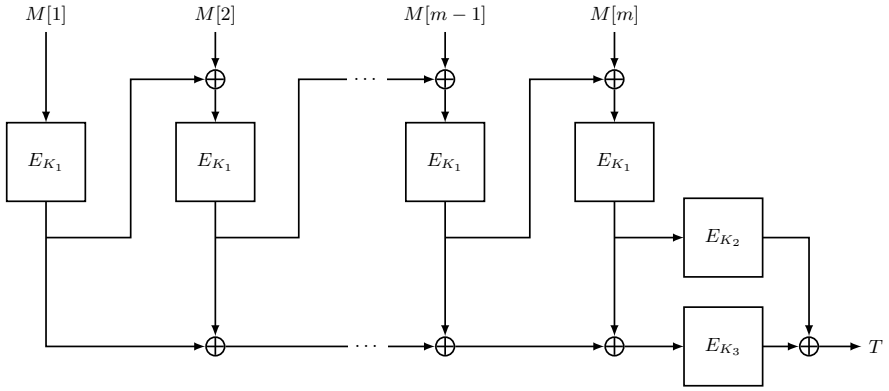


Fig. 5: 3kf9 based on a block cipher E using three keys K_1, K_2, K_3.

7 Security of 3kf9 and LightMAC-Plus

In this section, we provide upper bounds on the PRF-security of 3kf9 and LightMAC-Plus. Due to space constraints, the proof is deferred to the full version of this paper. We remark that the security proof of LightMAC-Plus is much simpler than PMAC-Plus; the structure of LightMAC-Plus is similar to PMAC-Plus, while domain separation by distinct prefixes removes most bad events in the proof.

7.1 Security of 3kf9

A $2n$-bit hash function 3kf9Hash is based an n-bit block cipher E using k-bit keys. For a padded message $M = M[1]\|M[2]\|\cdots\|M[m]$ where $m \leq \ell$, and for a key $K \in \{0,1\}^k$, $\mathsf{3kf9Hash}_K(M)$ is defined as follows.

Function $\mathsf{3kf9Hash}_K(M)$
  $Z[0] \leftarrow 0$
  for $\alpha \leftarrow 1$ to m **do**
    $Z[\alpha] \leftarrow E_K\left(Z[\alpha-1] \oplus M[\alpha]\right)$
  $U \leftarrow Z[m]$
  $V \leftarrow Z[1] \oplus Z[2] \oplus \cdots \oplus Z[m]$
  return (U, V)

The 3kf9 MAC is defined as DbHtS[3kf9Hash] (Figure 5). We prove the security of 3kf9 as follows.

Theorem 6. *Assume that $\ell \leq N/8$. When 3kf9 is based on a block cipher E, one has*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{3kf9}}(q,t,\ell) \leq \frac{18\ell^{\frac{4}{3}}q^{\frac{4}{3}}}{N} + \frac{2\ell^{\frac{2}{3}}q^{\frac{8}{3}}}{N^2} + \frac{2\ell^{\frac{4}{3}}q^{\frac{7}{3}}}{N^2} + \frac{11\ell^2q^2}{N^2} + \frac{11\ell^6q^4}{N^3} + 3\mathbf{Adv}^{\mathsf{prp}}_{E}(\ell q, t+t'),$$

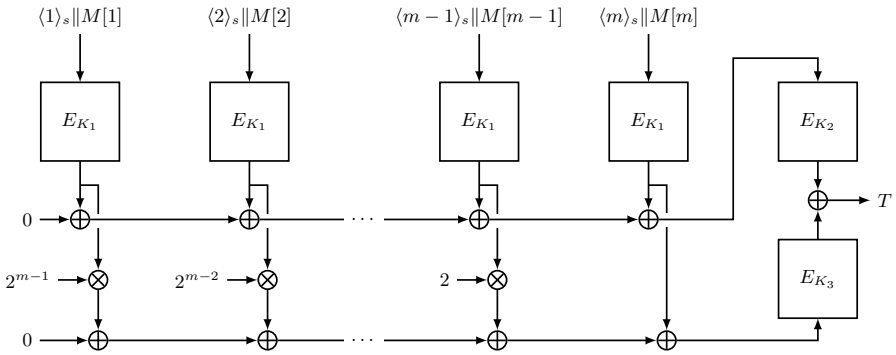


Fig. 6: LightMAC-Plus based on a block cipher E using three keys K_1, K_2, K_3.

where t' is the time complexity necessary to compute E for ℓq times.

7.2 Security of LightMAC-Plus

A $2n$-bit hash function LHash is based an n-bit block cipher E using k-bit keys. In this construction, a message is padded so that its length is a multiple of $n-s$, where s is a fixed parameter such that $0 < s < n$. So a padded message M can be broken into $(n-s)$-bit blocks; let

$$M = M[1]\|M[2]\|\cdots\|M[m],$$

where $m < 2^s$ and $M(\alpha)$ is $n-s$ bits for $\alpha = 1, \ldots, m$. Let $\langle\alpha\rangle_s$ denote the s-bit binary representation of integer α. Then for a key $K \in \{0,1\}^k$, $\mathsf{LHash}_K(M)$ is defined as follows.

Function $\mathsf{LHash}_K(M)$
  for $\alpha \leftarrow 1$ to m **do**
    $X[\alpha] \leftarrow \langle\alpha\rangle_s\|M[\alpha]$
    $Y[\alpha] \leftarrow E_K(X[\alpha])$
  $U \leftarrow Y[1] \oplus Y[2] \oplus \cdots \oplus Y[m]$
  $V \leftarrow 2^{m-1} \cdot Y[1] \oplus 2^{m-2} \cdot Y[2] \oplus \cdots \oplus Y[m]$
  return (U, V)

The LightMAC-Plus MAC is defined as DbHtS[LHash] (Figure 6). We prove the security of LightMAC-Plus as follows.

Theorem 7. *Assume that $\ell \le N/16$. When* LightMAC-Plus *is based on a block cipher E, one has*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{LightMAC\text{-}Plus}}(q,t,\ell) \le \frac{17q^{\frac{4}{3}}}{2N} + \frac{2}{N} + \frac{2\sqrt{2}q^2}{N^{\frac{3}{2}}} + \frac{3q^{\frac{8}{3}}}{N^2} + \frac{9q^{\frac{7}{3}}}{2N^2} + \frac{30q^2}{N^2} + \frac{44q^4}{3N^3} + 3\mathbf{Adv}^{\mathsf{prp}}_E(\ell q, t+t'),$$

where t' is the time complexity necessary to compute E for ℓq times.

References

[1] Subhadeep Banik, Sumit Kumar Pandey, Thomas Peyrin, Yu Sasaki, Siang Meng Sim, and Yosuke Todo. GIFT: A Small Present Towards Reaching the limit of Lightweight Encryption. In Wieland Fischer and Naofumi Homma, editors, *CHES 2017*, volume 10529 of *LNCS*, pages 321–345. Springer, 2017.

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

[3] 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.

[4] Jürgen Bierbrauer, Thomas Johansson, Gregory Kabatianskii, and Ben Smeets. On Families of Hash Functions via Geometric Codes and Concatenation. In Douglas R. Stinson, editor, *CRYPTO '93*, volume 773 of *LNCS*, pages 40–48. Springer, 1993.

[5] John Black and Phillip Rogaway. A Block-Cipher Mode of Operation for Parallelizable Message Authentication. In Lars R. Knudsen, editor, *EUROCRYPT 2002*, volume 2332 of *LNCS*, pages 384–397. Springer, 2002.

[6] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J.B. Robshaw, Yannick Seurin, and Charlotte Vikkelsoe. PRESENT: An Ultra-Lightweight Block Cipher. In Pascal Paillier and Ingrid Verbauwhede, editors, *CHES 2007*, volume 4727 of *LNCS*, pages 450–466. Springer, 2007.

[7] Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Goutam Paul. Double-block Hash-then-Sum: A Paradigm for Constructing BBB Secure PRF. *IACR Transactions on Symmetric Cryptology*, 2018(3):36–92, 2018.

[8] Bert den Boer. A Simple and Key-Economical Unconditional Authentication Scheme. *Journal of Computer Security*, 2:65–72, 1993.

[9] Jian Guo, Thomas Peyrin, Axel Poschmann, and Matt Robshaw. The LED block cipher. In Tsuyoshi Takagi Bart Preneel, editor, *CHES 2011*, volume 6917 of *LNCS*, pages 326–341. Springer, 2011.

[10] Tetsu Iwata and Kaoru Kurosawa. OMAC: One-Key CBC MAC. In Thomas Johansson, editor, *FSE 2003*, volume 2887 of *LNCS*, pages 129–153. Springer, 2003.

[11] Ashwin Jha and Mridul Nandi. Revisiting Structure Graphs: Applications to CBC-MAC and EMAC. *Journal of Mathematical Cryptology*, 10(3-4):157–180, 2016.

[12] Gaëtan Leurent, Mridul Nandi, and Ferdinand Sibleyras. Generic Attacks Against Beyond-Birthday-Bound MACs. In Hovav Shacham and Alexandra Boldyreva, editors, *Crypto 2018*, volume 10991 of *LNCS*, pages 306–336. Springer, 2018.

[13] Atul Luykx, Bart Preneel, Elmar Tischhauser, and Kan Yasuda. A MAC Mode for Lightweight Block Ciphers. In Thomas Peyrin, editor, *FSE 2016*, volume 9783 of *LNCS*, pages 43–59. Springer, 2016.

[14] Yusuke Naito. Blockcipher-based MACs: Beyond the Birthday Bound without Message Length. In Tsuyoshi Takagi and Thomas Peyrin, editors, *ASIACRYPT 2017*, volume 10626 of *LNCS*, pages 446–470. Springer, 2017.

[15] Jacques Patarin. Introduction to Mirror Theory: Analysis of Systems of Linear Equalities and Linear Non Equalities for Cryptography. IACR Cryptology ePrint Archive, Report 2010/287, 2010. Available at http://eprint.iacr.org/2010/287.

[16] Jacques Patarin. Mirror Theory and Cryptography. IACR Cryptology ePrint Archive, Report 2016/702, 2016. Available at http://eprint.iacr.org/2016/702.

[17] Richard Taylor. An Integrity Check Value Algorithm for Stream Ciphers. In Douglas R. Stinson, editor, *CRYPTO '93*, volume 773 of *LNCS*, pages 331–343. Springer, 1993.

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

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

[20] Liting Zhang, Wenling Wu, Han Sui, and Peng Wang. 3kf9: Enhancing 3GPP-MAC beyond the Birthday Bound. In Xiaoyun Wang and Kazue Sako, editors, *ASIACRYPT 2012*, volume 6841 of *LNCS*, pages 296–312. Springer, 2012.