IACR Transactions on Symmetric Cryptology
ISSN 2519-173X, Vol. 2017, No. 4, pp. 268–305. DOI:10.13154/tosc.v2017.i4.268-305

Single Key Variant of PMAC_Plus

Nilanjan Datta[1], Avijit Dutta[2], Mridul Nandi[2], Goutam Paul[2] and Liting Zhang[3]

[1] Indian Institute of Technology, Kharagpur, India
nilanjan_isi_jrf@yahoo.com
[2] Indian Statistical Institute, Kolkata, India
avirocks.dutta13@gmail.com,mridul.nandi@gmail.com,goutam.paul@isical.ac.in
[3] Westone Cryptologic Research Center, Beijing, China
liting.zhang@hotmail.com

Abstract. At CRYPTO 2011, Yasuda proposed the PMAC_Plus message authentication code based on an n-bit block cipher. Its design principle inherits the well known PMAC parallel network with a low additional cost. PMAC_Plus is a rate-1 construction like PMAC (i.e., one block cipher call per n-bit message block) but provides security against all adversaries (under black-box model) making queries altogether consisting of roughly upto $2^{2n/3}$ blocks (strings of n-bits). Even though PMAC_Plus gives higher security than the standard birthday bound security, with currently available best bound, it provides weaker security than PMAC for certain choices of adversaries. Moreover, unlike PMAC, PMAC_Plus operates with three independent block cipher keys. In this paper, we propose 1k-PMAC_Plus, the first rate-1 single keyed block cipher based BBB (Beyond Birthday Bound) secure (in standard model) deterministic MAC construction without arbitrary field multiplications. 1k-PMAC_Plus, as the name implies, is a simple one-key variant of PMAC_Plus. In addition to the key reduction, we obtain a higher security guarantee than what was proved originally for PMAC_Plus, thus an improvement in two directions.

Keywords: PMAC · PMAC_Plus · Beyond Birthday · Cover-free · PRF · Sum of PRPs.

1 Introduction

A Message Authentication Code (MAC) is a fundamental symmetric-key primitive that allows a sender to authenticate messages by computing tags that can be verified by the receiver holding a common secret key with the sender. In literature, there are several MACs which are based on block ciphers as fundamental primitives (e.g., CBC-MAC [BKR00], CMAC [NIS05], OMAC [IK03], GCBC [Nan09] etc). Among these, many block cipher based MACs are specified in a large number of standardized documents including ISO 9797-1 [JTC11]. Unlike these, PMAC (Parallelizable MAC) [BR02] is a distinctive, completely parallelizable block cipher based MAC. Under parallel implementation, PMAC outperforms CBC MACs significantly. Besides PMAC, there have been a few proposals of parallelizable block cipher based MACs, e.g. XOR MAC [BGR95], PCS [Ber99], LightMAC [LPTY16] etc. There is also some improvement over PMAC which includes the constructions PMAC1 [Rog04] and iPMAC [Sar10].

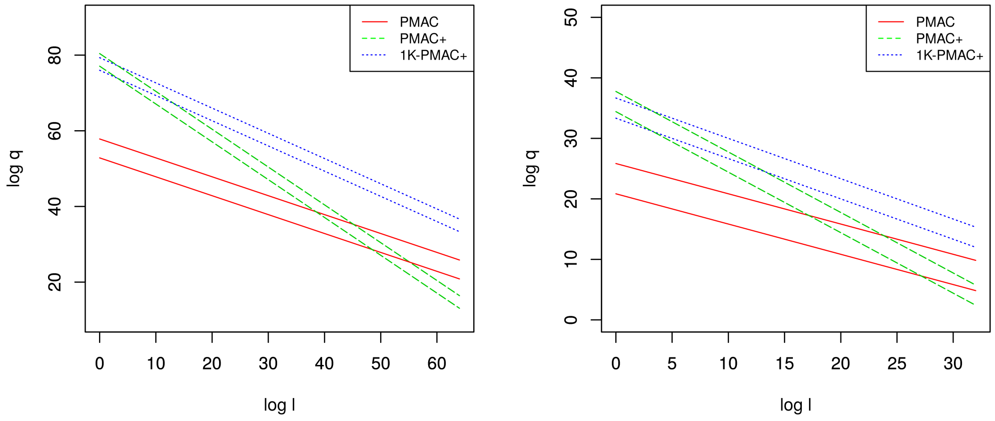
1.1 PMAC and PMAC_Plus

The main focus of this paper is around the design principle followed in PMAC and its pseudorandom function (PRF) security analysis. Informally, prf-advantage corresponds to the best advantage an adversary can achieve in distinguishing the concerned construction

Received: 2017-06-01, Revised: 2017-09-01, Accepted: 2017-11-01, Published: 2017-12-15

from a uniform random function (the ideal construction). Some known prf-advantages for PMAC are $\sigma^2/2^n$ [BR02], $10\ell q^2/2^n$ [MM07] and $5\sigma q/2^n$ [NM08] against all adversaries which are allowed to make at most q queries so that (i) total number of blocks in all queries is σ and (ii) the longest query contains at most ℓ blocks. Recently, Gaži et al. [GPR17] have shown that the bound $5\sigma q/2^n$ [NM08] is tight. These bounds [BR02, MM07, NM08] are called **birthday bounds** as the security bound becomes void after making roughly $2^{n/2}$ many queries. Yasuda in CRYPTO 2011 [Yas11], introduced a variant of PMAC, called PMAC_Plus, which achieves prf-advantages about $27\ell^3q^3/2^{2n}$. Even though the bound is beyond birthday in ℓq, we cannot conclude that PMAC_Plus always achieves higher security than PMAC as described below.

There are some choices of adversaries for which PMAC can provide a better security guarantee than the existing security guarantee of PMAC_Plus by Yasuda [Yas11]. Suppose, we have $n = 128$ and we want to fix the prf-advantage to be bounded by $\epsilon = 2^{-10}$. If the longest message consists of 2^{50} blocks, then PMAC permits about 2^{33} queries (using the bound $5\ell q^2/2^n$), whereas PMAC_Plus would permit queries fewer than 2^{31} queries (using the bound $27\ell^3q^3/2^{2n}$). Fig. 1.1 provides detail values of q for different choices of ℓ when the block length n is $128, 64$ bits with $\epsilon = 2^{-10}$ and $\epsilon = 2^{-20}$. PMAC_Plus also does not have improved bounds in terms of σ and q as we have for PMAC [NM08]. Suppose, we have only one large query consisting of 2^{50} blocks and the rest consisting of about 2^{20} blocks each, then roughly 2^{47} queries can be made for PMAC (using the improve bound $5\sigma q/2^n$). The bound given by PMAC_Plus does not give any advantage (permits less than 2^{31} queries as before) against such adversaries. Moreover, PMAC_Plus operates with three independent block cipher keys unlike PMAC which needs only one block cipher key. The author of PMAC_Plus has mentioned [Yas11] that it would be challenging to come up with a rate-1 single keyed block cipher based deterministic MAC with beyond birthday bound security. Similar challenges have been raised by authors of 3kf9 [ZWSW12] and



(a) $n = 128$

(b) $n = 64$

Figure 1.1: $\log \ell$ vs $\log q$ graph for PMAC, PMAC_Plus and 1k-PMAC_Plus construction with $\epsilon = 2^{-10}, 2^{-20}$. For a fixed construction the top curve is for $\epsilon = 2^{-10}$ and the bottom one is for $\epsilon = 2^{-20}$. Observe that, for $n = 128$, PMAC and PMAC_Plus intersects at $\ell = 2^{45.2}$ ($2^{48.5}$) for $\epsilon = 2^{-10}$ (2^{-20}). So, for any ℓ beyond that, PMAC always achieves better security than PMAC_Plus. Similarly for $n = 64$, the intersecting ℓ values are $2^{23.82}$ and $2^{27.15}$ resp.

EWCDM [CS16] which have beyond birthday bound security. *In this paper, we show the security bound of our proposed rate-*1 *single keyed block cipher based deterministic MAC construction* 1k-PMAC_Plus, *is beyond birthday secure and offers a better security guarantee than that of* PMAC_Plus.

1.2 Some Beyond Birthday Bound Constructions

Traditional schemes achieving BBB security either require the ideal cipher model [JJV02] or require a relatively large amount of randomness (at least $3n$ bits for the MACRX$_3$ construction of [BGK99]). MAC-R2 [Min10] as proposed by Minematsu in FSE 2010, uses a random n-bit IV but it is much slower. Nonce based MAC construction (e.g., EWCDM [CS16]) and random IV based MAC construction (MAC-R2) are more meaningful in the context of unforgeable security than PRF security. [1] Tweakable block cipher (TBC) based MAC constructions like PMAC2x [LN17], PMAC_TBC1K [Nai15] achieve optimal n bit security. These constructions are also very efficient as they require roughly one tweakable block cipher call per message block. However, we would like to mention that block ciphers (e.g. AES [DR00], DES [oS97]) are well studied, widely standardized and adopted primitive. In recent trend of cryptography, tweakable block ciphers (e.g. SKINNY [BJK$^+$16], Threefish [FLS$^+$10]) are also getting attention parallel to block ciphers. TBC with tweak size t bits and block size n bits potentially gives an $(n + t)/2$ bits of security if the input collisions are avoided. As tweakable block cipher can be viewed as an independent block cipher for each fixed setting of the tweak, it has overhead of processing tweak along with key as does in TWEAKEY framework [JNP14].

1.3 Our Contributions

The main contribution of the paper is to design a rate-1 single keyed (without generating multiple block cipher keys), block cipher based deterministic MAC construction with beyond birthday bound security. Clearly, one can derive multiple keys used in a construction by using some pseudorandom bit generator or using the underlying block cipher in a counter mode. However, there is no way to avoid key scheduling algorithms for multiple key based constructions. In this respect, 1k-PMAC_Plus, which is, to the best of our knowledge, the first rate-1 single keyed block cipher based beyond birthday bound secure deterministic MAC construction without arbitrary [2] field multiplications. We would like to mention that our proposed construction is very similar to the PMAC_Plus construction with minimal overhead cost. The notable features of 1k-PMAC_Plus are the following:

1. **Single Key with Minimal Cost and Overhead.** Unlike PMAC_Plus, 1k-PMAC_Plus requires a single block cipher key. Both constructions (i.e. PMAC_Plus and 1k-PMAC_Plus) require two masks and the masks can be derived from the underlying block cipher. Moreover, it is easy to see that a simple one key version (i.e. make all the three independent block cipher keys K_1, K_2 and K_3 as shown in Fig. 1.2, identical) of PMAC_Plus is clearly insecure as it returns zero output for any single block message. So, a modification on PMAC_Plus is required which ensures minimal cost and overhead. To achieve this, we multiply the intermediate value Θ_{old} (See Fig. 1.2) by the primitive element 2 of GF(2^n). We have also observed that xoring Θ_{old} by a non zero constant instead of multiplying it by 2, suffers from a birthday bound attack, as discussed in Sect. 4.1. Moreover, to get rid off

[1] Note that a simple nonce based construction, on an input message M and nonce N that returns $f_K(N)$, is a secure PRF where f is a PRF. Similar PRF construction based on random IV that ignores message input can be defined.

[2] By arbitrary, we mean any field multiplication except field multiplication by the primitive element 2 of $GF(2^n)$. As a matter of fact, field multiplication by 2 involves only shift and xor operations, which is cheap to implement in hardware.

the analysis of some extra bad events, we additionally introduce fix_0 and fix_1 function, as discussed in details in Sect. 4.

2. **Sum of Permutations Under Conditional Distribution.** We have shown that sum of two identical permutations over a restricted domain is a beyond birthday bound secure PRF. This result is a generalized version of *sum of two permutations* result, where the input space of permutations is a certain subset of the original input space. We require this generic result in the security analysis of the construction as the output of its last two block cipher calls do not have full entropy due to some previous assignments of block cipher outputs in the internal hash computation. Moreover, we believe this result to be handy in analysing the security of single-keyed block cipher based construction that inherently uses the sum function.

3. **Improved Security Bound.** We have obtained a $O(q\sigma^2/2^{2n})$ PRF security bound (also applicable to original the PMAC_Plus construction) for 1k-PMAC_Plus. Moreover, when all messages are of same length then the security bound of our construction becomes $O(q^3\ell^2/2^{2n})$ which compares favorably to that of $O(q^3\ell^3/2^{2n})$ as proved in the security bound for PMAC_Plus [Yas11]. This also ensures that 1k-PMAC_Plus always achieves higher security than PMAC (see Fig. 1.1). We would like to point out that our proven bound for 1k-PMAC_Plus also holds for PMAC_Plus and therefore the security bound of PMAC_Plus is improved upon its existing security bound.

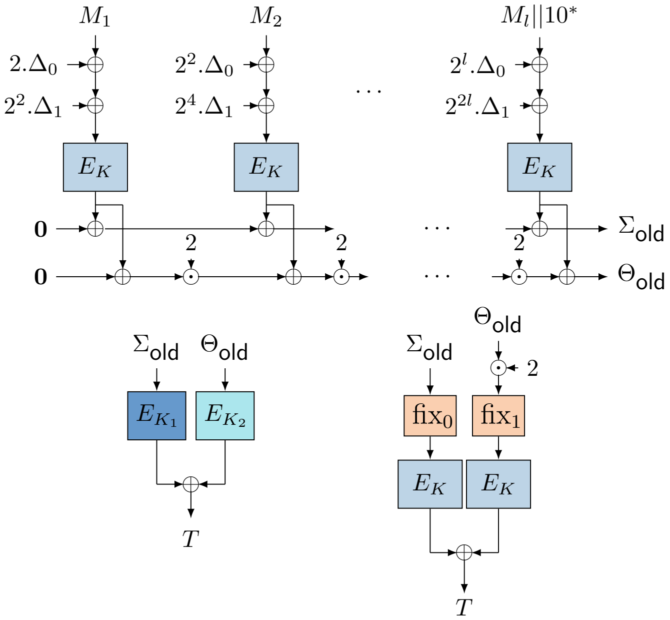


Figure 1.2: (i) Upper construction combined with the lower left construction is PMAC_Plus; (ii) Upper Construction combined with the lower right construction is our proposed construction 1k-PMAC_Plus. M_i is a n bit binary string, denotes the i-th message block and l denotes the number of message blocks. $\Delta_0 = E_K(\mathbf{0})$ and $\Delta_1 = E_K(\mathbf{1})$, where $\mathbf{0}$ is a n bit binary string consisting of all 0's and $\mathbf{1}$ is a n bit binary string consisting of all 0's with lsb set to 1. '2' is the primitive element of GF(2^n). fix_0 function takes n bit binary string as input value and returns the same input binary string with its lsb set to 0. Similarly, for fix_1 function, lsb is set to 1.

Table 1: BC denotes block cipher calls. Rate defines the average number of message blocks processed by a single execution of block cipher. l denotes the number of message blocks in a mesage, q denotes the total number of queries, ℓ denotes the maximum number of message blocks in all q queries and σ denotes the total number of message blocks in all q queries.

Construction	# of keys	BC	rate	Security Bound
PMAC [BR02]	1	$l+1$	1	$\frac{10\ell q^2}{2^n}$ [MM07], $\frac{5\sigma q}{2^n}$ [NM08]
SUM-ECBC [Yas10]	4	$2(l+1)$	1/2	$\frac{40\ell^3 q^3}{2^{2n}}$
PMAC_Plus [Yas11]	3	$l+2$	1	$\frac{27q^3\ell^3}{2^{2n}}$
3kf9 [ZWSW12]	3	$l+2$	1	$\frac{4q^3\ell^3}{2^{2n}} + \frac{4q\ell}{2^n}$
1k-PMAC_Plus [This Paper]	1	$l+2$	1	$\frac{21\sigma}{2^n} + \frac{224q\sigma^2}{2^{2n}}$

2 Preliminaries

2.1 Symbol and Notation

We fix a positive integer n and write $N = 2^n$. An element of $\mathbb{B} := \{0,1\}^n$ is said to be a **block** which is a bit string of length n, where n denotes the block length which is typically 64 or 128 bits. Let $\mathrm{GF}(2^n)$ be the field of order 2^n. We identify bit strings and finite field elements of $\mathrm{GF}(2^n)$ by representing the string $a = a_{n-1}a_{n-2}\ldots a_1a_0 \in \mathbb{B}$ as polynomial $a(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1x + a_0 \in \mathrm{GF}(2^n)$ and vice versa. For $a, b \in \{0,1\}^n$, we define field addition $a \oplus b$ as addition of the polynomials $a(x) + b(x) \in \mathrm{GF}(2^n)$. Multiplication $a \odot b$ is defined with respect to the irreducible polynomial $f(x)$ used to represent $\mathrm{GF}(2^n)$ as $a(x) \cdot b(x) \mod f(x)$. Therefore, we can view $\mathbb{B}$ as the finite field $\mathrm{GF}(2^n)$ with $\oplus$ as field addition and $\odot$ as field multiplication. $\{0,1\}^* := \cup_{i \geq 0}\{0,1\}^i$ denotes the set of all possible binary strings of arbitrary length. We write $\mathbf{0}$ and $\mathbf{1}$ to denote the binary string 0^n and $0^{n-1}\|1$ respectively. For $a, b \in \mathbb{B}$, we write $a =_1 b$ to denote $a \in \{b, b \oplus \mathbf{1}\}$. In other words, $a =_1 b$ gives either of the following two equalities: (i) $a = b$ or (ii) $a = b \oplus \mathbf{1}$ but not both. For a given ordered set $\mathcal{S}$ we write $\min \mathcal{S}$ and $\min_2 \mathcal{S}$ to denote the minimum and second minimum element of $\mathcal{S}$ respectively.

The set of all functions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted as $\mathsf{Func}(\mathcal{X}, \mathcal{Y})$ and the set of all permutations over $\mathcal{X}$ is denoted as $\mathsf{Perm}(\mathcal{X})$. A function f mapping an element from arbitrary domain to $\mathbb{B}$ is called a **block function**. Similarly, a permutation over $\mathbb{B}$ is called the **block permutation**. The set of all block functions with domain $\mathcal{X}$ is denoted as $\mathsf{Func}_{\mathcal{X}}$ and the set of all block permutations is denoted as Perm. We often write $\mathsf{Func}_{\mathcal{X}}$ as Func when the domain of the functions (i.e. $\mathcal{X}$) is understood from the context.

We denote a tuple $x := (x^i : i \in I)$ over an index set I as x^I to emphasize the index set I. We use the notation $(x^i)_i$ to denote a tuple when the index set I is clear from the context. An element x^i of a tuple x could itself be a tuple (in this paper, context wise we consider an element x^i of a tuple x is a tuple of size 2, i.e. $x^i := (x_0^i, x_1^i)$ and thus in that case we denote the tuple x as $(x_0^i, x_1^i)_i$). A natural choice of index set that we often use in the paper is $[q] := [1..q] := \{1, 2, \ldots, q\}$ for a positive integer q. A tuple $(x^i)_i$ is called a **block tuple** if every element of the tuple is a member of the set $\mathbb{B}$. Number of elements x^i of a tuple x^I is called the size of the tuple, denoted by $||x^I||$. Union of two tuples x and y is denoted by $x \cup y$. Similarly, we denote the intersection of two sets $\mathcal{X}$ and $\mathcal{Y}$ as $\mathcal{X} \cap \mathcal{Y}$. If $\mathcal{X} \cap \mathcal{Y} = \emptyset$ then we use the notation $\mathcal{X} \sqcup \mathcal{Y}$ to represent the disjoint union. An element x^i is said to be **fresh** in a tuple x if for all $j \neq i$, $x^i \neq x^j$. Otherwise we call the element to be **non-fresh** in tuple x. We call a pair of block tuple (x, y) to be **permutation compatible**, if there exists a permutation $\pi \in \mathsf{Perm}$ such that $\pi(x^i) = y^i$

where $x := (x^i : i \in I)$ and $y := (y^i : i \in I)$.

For two integers a, b such that $a \geq b$, we use the notation $\mathbf{P}(a, b) := \prod_{i=1}^{b}(a - (i-1))$ to denote the number of permutations of a distinct objects taken b at a time. For any set $\mathcal{B}$, we write $\mathcal{B}^{(s)} = \{(x^1, \ldots, x^s) \in \mathcal{B}^s : \forall i \neq j,\ x^i \neq x^j\}$. If $|\mathcal{B}| = m$ then $|\mathcal{B}^{(s)}| = \mathbf{P}(m, s)$.

We denote $X \leftarrow_{\$} S$ to mean that X is sampled uniformly at random from a finite set S and independently to all other random variables defined so far. Similarly, we denote $X_1, \ldots, X_i \xleftarrow{\text{wor}} S$ to mean that $X_1, \ldots, X_i$ are sampled without replacement from a finite set S.

2.2 Security Notion

Let $\mathcal{A}$ be an oracle algorithm that has access to its oracle $\mathcal{O}$. It makes finitely many queries adaptively to its oracle $\mathcal{O}$ and after the interaction it outputs a bit which we denote as $\mathcal{A}^{\mathcal{O}(\cdot)}$. We denote the **transcript** of $\mathcal{A}$ by the pair $(x^{[s]} := (x^1, \ldots, x^s), y^{[s]} := (y^1, \ldots, y^s)$ where x^i is the i-th query by $\mathcal{A}$ to oracle $\mathcal{O}$ and y^i be the corresponding response of $\mathcal{O}$. Given an oracle adversary $\mathcal{A}$, we define the prf-advantage of $\mathcal{A}$ against a keyed function family F over a domain $\mathcal{D}$ that outputs n bits as

$$\mathbf{Adv}_F^{\text{prf}}(\mathcal{A}) := |\Pr[\mathcal{A}^{F_K} = 1 : K \leftarrow_{\$} \mathcal{K}] - \Pr[\mathcal{A}^{\rho} = 1; \rho \leftarrow_{\$} \mathsf{Func}_{\mathcal{D}}]|.$$

Similarly, we define prp-advantage of $\mathcal{A}$ against a keyed function family F that output n bits as

$$\mathbf{Adv}_F^{\text{prp}}(\mathcal{A}) := |\Pr[\mathcal{A}^{F_K} = 1; K \leftarrow_{\$} \mathcal{K}] - \Pr[\mathcal{A}^{\Pi} = 1; \Pi \leftarrow_{\$} \mathsf{Perm}]|.$$

In the above definition of advantage function the probabilities are defined over internal coin tosses of $\mathcal{A}$ (if any) and the choice of K and ρ (or Π) depending on prf (or prp)-advantage. $\mathbf{Adv}_F^{\text{xxx}}(q, \sigma, t)$ [3] denotes $\max_{\mathcal{A}} \mathbf{Adv}_F^{\text{xxx}}(\mathcal{A})$ where xxx is either prf or prp and maximum is taken over all adversaries $\mathcal{A}$ running in time t, making at most q queries with an aggregate of total σ blocks.

2.3 Coefficients H Technique

In this section, we briefly discuss Coefficients H Technique [Pat08a] due to Patarin. Let us consider a distinguisher $\mathcal{A}$ with access to an oracle $\mathcal{O}$ and we assume that $\mathcal{A}$ is deterministic. When $\mathcal{A}$ interacts with $\mathcal{O}$, it issues queries to the oracle and obtains response from it. After this interaction is over, $\mathcal{A}$ outputs a decision bit. The collection of all queries and responses that is made to and from the oracle during the interaction of $\mathcal{A}$ with $\mathcal{O}$, is called a **transcript** of $\mathcal{A}$, denoted as $\tau^{\mathcal{A}}$. For the sake of simplicity of analysis, we slightly modify the experiment where we release internal information about the oracle to $\mathcal{A}$ only after $\mathcal{A}$ completes all queries and responses but before it outputs its decision. That is, we are making the distinguisher $\mathcal{A}$ more powerful by releasing extra information about the oracle. In this case, $\tau^{\mathcal{A}}$ contains the additional information and clearly the distinguishing advantage of $\mathcal{A}$ in the modified experiment can not be less than the distinguishing advantage of $\mathcal{A}$ in the former one.

Let X_{re} (resp. X_{id}) denotes the random variable representing real world ($\mathcal{O}_{\text{re}}$) and ideal world ($\mathcal{O}_{\text{id}}$) transcript respectively. The probability of realizing a transcript τ in ideal world (i.e. $\Pr[X_{\text{id}} = \tau]$) is called **ideal interpolation probability**. Similarly, the probability of realizing a transcript τ in real world (i.e $\Pr[X_{\text{re}} = \tau]$) is called **real interpolation probability**. A transcript τ is said to be **attainable** with respect to $\mathcal{A}$ if the ideal interpolation probability is non zero (i.e. $\Pr[X_{\text{id}} = \tau] > 0$). We denote the set of all attainable transcripts by $\mathcal{V}$. Following these notations, we state the main theorem of Coefficients H Technique as follows.

[3] Sometimes, the resources are also measured in terms of q, ℓ and t, where ℓ denotes the maximum number of blocks in a message.

Theorem 1 (Coefficients H Technique). *Let $\mathcal{V} = \mathcal{V}_{\text{good}} \sqcup \mathcal{V}_{\text{bad}}$ be some partition of the set of attainable transcripts. Suppose there exists $\epsilon_{\text{ratio}} \geq 0$ such that for any $\tau \in \mathcal{V}_{\text{good}}$,*

$$\frac{\Pr[X_{\text{re}} = \tau]}{\Pr[X_{\text{id}} = \tau]} := \frac{\mathsf{ip}_{\text{real}}}{\mathsf{ip}_{\text{ideal}}} \geq 1 - \epsilon_{\text{ratio}},$$

and there exists $\epsilon_{\text{bad}} \geq 0$ such that $\Pr[X_{\text{id}} \in \mathcal{V}_{\text{bad}}] \leq \epsilon_{\text{bad}}$. Then,

$$\mathbf{Adv}^{\mathcal{O}_{\text{id}}}_{\mathcal{O}_{\text{re}}}(\mathcal{A}) \leq \epsilon_{\text{ratio}} + \epsilon_{\text{bad}}. \tag{1}$$

When $\mathcal{O}_{\text{id}}$ is a function chosen uniformly at random and $\mathcal{O}_{\text{re}}$ is some keyed construction in our interest defined over the same domain, then Eqn. (1) says that $\mathbf{Adv}^{\text{prf}}_{\mathcal{O}_{\text{re}}}(\mathcal{A}) \leq \epsilon_{\text{ratio}} + \epsilon_{\text{bad}}$.

2.4 Basic Results of Linear Algebra

For any $a \times s$ matrix A, $A[i][j]$ denotes the element in the i-th row and j-th column of A. If $a \leq s$, we use the notation $A[\cdot, i..(i + a - 1)]$ to denote a square submatrix of A containing the columns $i, i + 1, \ldots, i + a - 1$ and all the rows. Given a column vector c of dimension $a \times 1$, we write $(A : c)$ to denote a new matrix formed by appending the vector c to A. This new matrix is called "**augmented matrix**", has dimension $a \times (s + 1)$. It is easy to see that $rank(A) \leq \min\{a, s\}$. For any row vector $Y := (Y_1, \ldots, Y_s)$ of dimension $1 \times s$, Y^{T} denotes the following column vector

$$Y^{\mathsf{T}} := \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \end{pmatrix}$$

of dimension $s \times 1$. Y^{T} is called the **transpose** of the row vector Y.

Consider a system of linear equations of the form $A \cdot Y^{\mathsf{T}} = c$ over unknown variables $Y_1, \ldots, Y_s$, where each element of A, Y and c are elements of the field $\mathbb{B}$. This system of linear equations is said to be **consistent** if it has at least one solution, otherwise we call it **inconsistent**. Note that a system of linear equations is consistent if and only if $rank(A) = rank(A : c)$. It has a unique solution if $rank(A) = s$ and it has many solutions if $rank(A) < s$.

For a given system of equations $A \cdot Y^{\mathsf{T}}$, where A is a matrix of dimension $a \times s$ with rank r and all it's elements are members of the field $\mathbb{B}$ and $Y := (Y_1, \ldots, Y_s) \xleftarrow{\text{wor}} \mathcal{S} \subseteq \mathbb{B}$ with $|\mathcal{S}| = N'$, the probability of realizing of a particular solution is at most $\frac{1}{\mathbf{P}(N' - s + r, r)}$ as stated formally in the following proposition.

Proposition 1. *Let $Y := (Y_1, \ldots, Y_s) \xleftarrow{\text{wor}} \mathcal{S} \subseteq \mathbb{B}$ where $|\mathcal{S}| = N'$ and A be a given matrix of dimension $a \times s$. Then, for any given column vector c of dimension $a \times 1$, we have*

$$\Pr[(A)_{a \times s} \cdot Y^{\mathsf{T}} = c] \leq \frac{1}{\mathbf{P}(N' - s + r, r)},$$

where r is the rank of the matrix A.

Observe that the number of ways we can choose the coefficients of the non-basis vectors is at most $\mathbf{P}(N', s - r)$ which uniquely determines the coefficient of the basis vectors. Moreover, the number of ways we can choose s-many variables is $\mathbf{P}(N', s)$. Dividing the former by later gives the required probability bound.

The following corollary follows from Proposition 1:

Corollary 1. *Let α, β and γ are three non-zero distinct constants such that $\alpha, \beta, \gamma \in [1, N-1]$. Let $\Delta_0, \Delta_1 \xleftarrow{\text{wor}} \mathbb{B}$. Then, for any $c_1, c_2 \in \mathbb{B}$ we have,*

$$(a) \quad \Pr[2^\alpha \Delta_0 \oplus 2^\beta \Delta_1 \oplus c_1 = \mathbf{0}] \leq \frac{1}{N-1},$$

$$(b) \quad \Pr[2^\alpha \Delta_0 \oplus 2^\beta \Delta_1 \oplus c_1 = \mathbf{0}, 2^\alpha \Delta_0 \oplus 2^\gamma \Delta_1 \oplus c_2 = \mathbf{0}] \leq \frac{1}{(N-1)(N-2)},$$

where 2 is the primitive element of $\mathrm{GF}(2^n)$.

3 Sum of Identical Random Permutation Under Conditional Distribution

In this section, we discuss a simple variant of sum of random permutations result: sum of two identical random permutation Π is a beyond birthday secure pseudorandom function even if we restrict some of the inputs and outputs of the random permutation Π. Sum of PRP is a popular approach for constructing a PRF from two PRPs [BGK99, BI99, Luc00, Pat08b, Pat10, CLP14]. In [BI99, Pat10, DHT17] the optimal security of the construction has been shown. However, none of the above works considered sum of the two identical random permutation under conditional distribution. In this section, we show that the sum of two identical random permutation under conditional distribution is a beyond birthday secure pseudorandom function.

Let $X = \{x_1, \ldots, x_s\}$ and $Y = \{y_1, \ldots, y_s\}$ be two sets where elements of each set is a member of $\mathbb{B}$. It is well known that, for any random permutation Π, if we condition on the event $\Pi(x_i) = y_i$, $1 \leq i \leq s$ then for any $x \in \mathbb{B} \setminus \{x_1, \ldots, x_s\}$, the conditional random variable $\Pi(x) \mid (\Pi(x_i) = y_i, \forall i \in [s])$ is distributed uniformly on the set $\mathbb{B} \setminus Y$. *In other words, the conditional distribution of a random permutation is same as random bijective function with an appropriate domain and range.* A random bijective function is a bijective function sampled uniformly at random from the set of all bijective functions with appropriate domain and range. The following result informally says that, sum of bijective functions also behaves close to a random function. We believe that the following result could be useful whenever we study a construction based on a single random permutation that involves sum function implicitly.

Theorem 2 (Sum of Identical Random Permutation Under Conditional Distribution)**.** *For any set Y of size s and a r tuple $t^{[r]} := (t^1, \ldots, t^r)$ of* non zero *n bit strings, let*

$$\mathcal{H} = \{(h_0^i, h_1^i)_i : h_0^i \oplus h_1^i = t^i \ \forall i \in [r],\ (h_0^i, h_1^i)_i \in (\mathbb{B} \setminus Y)^{(2r)}\}.$$

Then, $|\mathcal{H}| \geq \frac{\mathbf{P}(N-s,2r)}{N^r}(1-\mu_2)$ where $\mu_2 = \frac{rs^2+2sr^2+4r^3/3}{(N-s-2r)^2}$. Moreover, if $s+2r \leq \frac{N}{2}$, then $\mu_2 \leq \frac{4rs^2+8sr^2+6r^3}{N^2}$.

Proof. For each $j \in [r]$, we define the following set

$$\mathcal{H}_j := \{(h_0^i, h_1^i)_i : h_0^i \oplus h_1^i = t^i \ \forall i \in [j],\ (h_0^i, h_1^i)_i \in (\mathbb{B} \setminus Y)^{(2j)}\}.$$

Note that

$$|\mathcal{H}_j| \geq |\mathcal{H}_{j-1}| \times |\{(h_0^j, h_1^j) : h_0^j \oplus h_1^j = t^j,\ (h_0^j, h_1^j) \in (\mathbb{B} \setminus Y_j)^2\}|, \tag{2}$$

where $Y_j = Y \sqcup \{a_1, \ldots, a_{2(j-1)}\}$ such that $a_{2i-1} \oplus a_{2i} = t^i$, $\forall i \in [j-1]$. Let, $s' := |Y_j| = s + 2(j-1)$. Now, we make the following claim, the proof of which is postponed to later in the section.

Claim 1. For any fixed $j \in [r]$, the cardinality of the set

$$\mathcal{T}_j := \{(h_0^j, h_1^j) : h_0^j \oplus h_1^j = t^j (\neq \mathbf{0}),\ (h_0^j, h_1^j) \in (\mathbb{B} \setminus Y_j)^2\},$$

where $Y_j = Y \sqcup \{a_1, \ldots, a_{2(j-1)}\}$ such that $a_{2i-1} \oplus a_{2i} = t^i (\neq \mathbf{0})$, $\forall i \in [j-1]$, is given by

$$|\mathcal{T}_j| \geq \frac{(N-s')(N-s'-1)}{N} \cdot \Big(1 - \frac{s'^2}{(N-s')^2}\Big).$$

Now, we resume our proof. From Eqn. (2) and Claim 1, we write the following:

$$|\mathcal{H}_j| \geq |\mathcal{H}_{j-1}| \times \frac{(N-s')(N-s'-1)}{N} \cdot \Big(1 - \frac{s'^2}{(N-s')^2}\Big).$$

A simple algebraic calculation yields the following lower bound on $|\mathcal{H}_j|$

$$|\mathcal{H}_j| \geq \prod_{i=0}^{j-1} \frac{\big(N-(s+2i)\big) \cdot \big(N-(s+2i)-1\big)}{N} \cdot \Big(1 - \epsilon_{i+1}\Big), \tag{3}$$

where $\epsilon_{i+1} := \frac{(s+2i)^2}{\big(N-(s+2i)\big)^2}$. Now, we calculate the bound on the cardinality of $\mathcal{H}_r$. Note that by definition $\mathcal{H}_r = \mathcal{H}$. Therefore, a bound on $|\mathcal{H}_r|$ is sufficient. It is easy to see from Eqn. (3) that $|\mathcal{H}_r| \geq \frac{\mathbf{P}(N-s,2r)}{N^r} \cdot (1 - \sum_{i=1}^{r} \epsilon_i)$, where one can easily check that $\sum_{i=1}^{r} \epsilon_i \leq \frac{rs^2+2sr^2+4r^3/3}{(N-s-2r)^2} (= \mu_2)$. Moreover, it is easy to see that, if $s + 2r \leq \frac{N}{2}$, then $\mu_2 \leq \frac{4rs^2+8sr^2+6r^3}{N^2}$.

Proof of Claim 1. In the proof of this claim, our primary interest is to obtain a lower bound on $|\mathcal{T}_j|$. To do this, let us define two more sets: for each $b = 0, 1$, $\mathcal{T}_j^b := \{(h_0^j, h_1^j) : h_0^j \oplus h_1^j = t^j,\ h_b^j \in Y_j\}$. Clearly, for each $b = 0, 1$, we have $|\mathcal{T}_j^b| \leq s'$, where $|Y_j| = s + 2(j-1) (= s')$. Therefore,

$$\begin{aligned} |\mathcal{T}_j| &\geq N - |\mathcal{T}_j^0 \cup \mathcal{T}_j^1| \\ &\geq N - 2s' \\ &\geq \frac{(N-s')(N-s'-1)}{N} \cdot \left(1 - \frac{s'^2}{(N-s')^2}\right). \end{aligned}$$

Hence, our result follows.

4 1k-PMAC_Plus: Design and Security Claim

Our construction of 1k-PMAC_Plus is shown in Fig. 4.1 and the algorithmic description is given in Fig. 4.2. As shown in Fig. 4.2, 1k-PMAC_Plus differs from the existing PMAC_Plus construction in one extra field multiplication by 2 (which is nothing but shift and xor operation), in applying the fix_0, fix_1 functions and obviously replacing three independent block cipher keys with the same block cipher key. Recall that, for $b \in \{0, 1\}$, fix_b is a function takes an n-bit binary string as input and returns an n-bit binary string with least significant bit fixed to bit b, keeping all other remaining bits same (i.e. $\mathsf{fix}_0(x_1, \ldots, x_n) = x_1, \ldots, x_{n-1}0$). As a matter of fact, we have used the notation fix_0 and fix_1 to separate the range of collision of Σ, Θ, which reduces the analysis of some bad cases (e.g. Σ^i cannot be equal to Θ^j for some i, j) and simplies the proof.

In this paper, we show that 1K-PMAC_Plus[Π] (1k-PMAC_Plus instantiated with random permutation Π) is indistinguishable from random function up to roughly $2^{2n/3}$ message blocks. More formally, we state the following security result about 1k-PMAC_Plus[E] (1k-PMAC_Plus instantiated with a block cipher E).

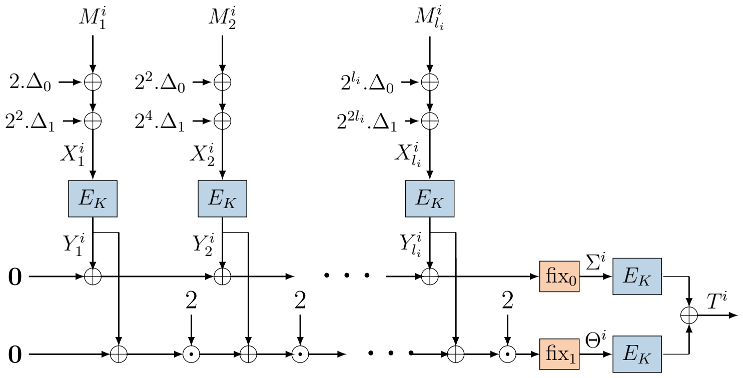


Figure 4.1: 1k-PMAC_Plus Construction.

Theorem 3. *Any distinguisher with running time t, making q-tuple of distinct messages with an aggregate of total σ-many blocks, can distinguish* 1k-PMAC_Plus$[E]$ *from a uniform random function by,*

$$\mathbf{Adv}^{\text{prf}}_{\text{1k-PMAC\_Plus}[E]}(q,\sigma,t) \leq \mathbf{Adv}^{\text{prp}}_{E}(\sigma+2+2q,t') + \frac{21\sigma}{N} + \frac{224q\sigma^2}{N^2},$$

where $t' = t + O(\sigma + 2q + 2)$.

```
Internal$[E_K](M)$
    $\Delta_0 \leftarrow E_K(0)$
    $\Delta_1 \leftarrow E_K(1)$
    $M_1\|\ldots\|M_l \twoheadleftarrow M\|10^*$
    for $j = 1$ to $l$:
        $X_j \leftarrow M_j \oplus 2^j\Delta_0 \oplus 2^{2j}\Delta_1$
        $Y_j \leftarrow E_K(X_j)$
    $\Sigma_{\text{old}} \leftarrow Y_1 \oplus Y_2 \oplus \ldots \oplus Y_l$
    $\Theta_{\text{old}} \leftarrow 2^{l-1}\cdot Y_1 \oplus 2^{l-1}\cdot Y_2 \oplus \ldots \oplus Y_l$
    return $(\Sigma_{\text{old}}, \Theta_{\text{old}})$
```

```
PMAC_Plus$[E_{K,K_1,K_2}](M)$
    $(\Sigma_{\text{old}}, \Theta_{\text{old}}) \leftarrow$ Internal$[E_K](M)$
    $T = E_{K_1}(\Sigma_{\text{old}}) \oplus E_{K_2}(\Theta_{\text{old}})$
    return $T$

1K-PMAC_Plus$[E_K](M)$
    $(\Sigma_{\text{old}}, \Theta_{\text{old}}) \leftarrow$ Internal$[E_K](M)$
    $\Sigma \leftarrow \text{fix}_0(\Sigma_{\text{old}})$
    $\Theta \leftarrow \text{fix}_1(2\cdot\Theta_{\text{old}})$
    $T = E_K(\Sigma) \oplus E_K(\Theta)$
    return $T$
```

Figure 4.2: 1K-PMAC_Plus Construction; $M\|10^*$ denotes that 10^* string is padded to M to make the size (in number of bits) of the total message multiple of n. $M_1\|\ldots\|M_l \twoheadleftarrow M\|10^*$ denotes $M\|10^*$ is partitioned into l-many blocks each of length n bits where n is the block length of the underlying block cipher.

4.1 Design and Specification of 1k-PMAC_Plus

Let us first see whether the existing construction of PMAC_Plus is secure or not in the single key setting (meaning $K_1 = K_2 = K$). We observe that PMAC_Plus has a very trivial attack in the single key setting as querying a single block message would make Σ_{old} and Θ_{old} identical which gives the output $\mathbf{0}$ with probability 1. So, we look for a modified version of PMAC_Plus with minimal changes.

The possible minimal changes on PMAC_Plus in this direction are:

1. xoring a non-zero constant c with Θ_{old} (i.e. $\Theta = \Theta_{\mathsf{old}} \oplus c$), and
2. multiplying a primitive element 2 with Θ_{old} (i.e. $\Theta = 2 \cdot \Theta_{\mathsf{old}}$).

We observe that first modified construction has the following birthday bound attack, and hence we opt for the second choice.

4.2 Failure Attempt

In this section, we discuss why xoring a non-zero constant c with Θ_{old} does not yield beyond birthday bound security. Consider a distinguisher $\mathcal{A}$ that makes distinct single block message queries $m^1, \ldots, m^{\sqrt{N}}$. Suppose the corresponding outputs be $T^1, \ldots, T^{\sqrt{N}}$. If $\exists i \neq j$ such that $T^i = T^j$, $\mathcal{A}$ returns 1, o.w. 0.

Let us call the event $\exists i \neq j$ such that $T^i = T^j$ as COLL_T. Define $Z_1^i := E_K(\Sigma_{\mathsf{old}}^i) = E_K(Y^i)$ and $Z_2^i := E_K(\Theta^i) = E_K(Y^i \oplus c)$. So, $T^i = Z_1^i \oplus Z_2^i$ and $T^j = Z_1^j \oplus Z_2^j$.

It is easy to see, that the probability of holding the event COLL_T in ideal oracle is upper bounded by $\frac{\sqrt{N}(\sqrt{N}-1)}{2N}$.

Now in real world, the probability of holding the event COLL_T is calculated as:

$$\begin{aligned}
\Pr[\mathsf{COLL}_T] &= \sum_{i,j} \Pr[T^i = T^j \wedge Y^i \oplus Y^j = c] + \Pr[T^i = T^j \wedge Y^i \oplus Y^j \neq c] \\
&\overset{(1)}{=} \sum_{i,j} \Pr[T^i = T^j \mid Y^i \oplus Y^j = c] \cdot \Pr[Y^i \oplus Y^j = c] \\
&\quad\quad + \Pr[T^i = T^j \mid Y^i \oplus Y^j \neq c] \cdot \Pr[Y^i \oplus Y^j \neq c] \\
&\overset{(2)}{=} \sum_{i,j} 1 \cdot \frac{1}{N - \sqrt{N} + 1} + \frac{1}{N - \sqrt{N} + 1} \cdot \left(1 - \frac{1}{N - \sqrt{N} + 1}\right) \\
&\overset{(3)}{\geq} \sum_{i,j} \frac{2}{N} - \frac{1}{(N - \sqrt{N} + 1)^2} = \frac{\sqrt{N}(\sqrt{N} - 1)}{N} - \frac{\sqrt{N}(\sqrt{N} - 1)}{2(N - \sqrt{N} + 1)^2}.
\end{aligned}$$

Here we have used the following simple facts:

- By definition, $Y^i \oplus Y^j = c$ implies $T^i = T^j$.
- Y^i and Y^j are two wor samples over a set of size $(N - \sqrt{N} - 2)$. Hence,

$$\Pr[Y^i \oplus Y^j = c] = \frac{1}{(N - \sqrt{N} - 2) - 1} = \frac{1}{N - \sqrt{N} + 1}.$$

- $Y^i \oplus Y^j \neq c$ implies that Z_1^i, Z_2^i, Z_1^j and Z_2^j are all distinct, and hence wor samples over a set of size $(N - \sqrt{N} - 4)$. So, the event $T^i = T^j$ given the event $Y^i \oplus Y^j \neq c$, means distinct Z_1^i, Z_2^i, Z_1^j such that $Z_1^i \oplus Z_2^i \oplus Z_1^j \oplus Z_2^j = 0$. Therefore,

$$\Pr[T^i = T^j \mid Y_i \oplus Y_j \neq c] = \frac{1}{N - (\sqrt{N} - 4) - 3} = \frac{1}{N - \sqrt{N} + 1}.$$

Therefore, the advantage of $\mathcal{A}$ is given as

$$\mathbf{Adv}(\mathcal{A}) \geq \frac{\sqrt{N}(\sqrt{N}-1)}{2N} - \frac{\sqrt{N}(\sqrt{N}-1)}{2(N-\sqrt{N}+1)^2} \geq \frac{1}{2} - \frac{1}{2\sqrt{N}} - \frac{N}{2(N-\sqrt{N}+1)^2} \approx \frac{1}{2}.$$

5 Security Analysis of 1k-PMAC_Plus

In this section, we prove Theorem 3 using the Coefficients H technique. Before going into the details of the proof, we would like to give a brief overview of the proof in the following section.

5.1 Proof Idea of 1k-PMAC_Plus

In this section, we provide a brief proof sketch of the security theorem of 1k-PMAC_Plus. We extend the main proof idea of PMAC_Plus to the single-key setting and use the Coefficients H Technique to bound the PRF advantage of it. Before that, we define the following, which will help us to understand the proof idea.

Definition 1 (**Extended Cover-Free Tuple).** A tuple $(\Sigma^{[q]}, \Theta^{[q]})$ is said to be an extended covered tuple if $\exists i \in [q]$ such that Σ^i is non-fresh in $\Sigma^{[q]} \cup X$ and Θ^i is non-fresh in $\Theta^{[q]} \cup X$, where $X := (X^i_j : i \in [q], j \in [l_i])$ denotes the input tuple of internal hash computation (see Internal algorithm in Fig. 4.2). If no such i exists then the tuple is said to be an extended cover-free (e.c.f) tuple. [4]

We fix a q-tuple output $T^{[q]} := (T^1, \ldots, T^q)$ such that each $T^i \neq \mathbf{0}$. We identify some bad events and claim that if the bad events do not happen then the output distribution of 1k-PMAC_Plus is indistinguishable from uniform distribution close to upto $2^{2n/3}$ many blocks. Thus, to obtain the security bound for 1k-PMAC_Plus, we bound the following bad events in the ideal world:

5.1.1 Σ^i and Θ^i are non-fresh

If both Σ^i and Θ^i are non-fresh, then the bad event ECF occurs (defined in Fig. 5.1). In case of PMAC_Plus, if both Σ^i_{old} and Θ^i_{old} are non-fresh, then $\exists j, k < i$ such that $\Sigma^i_{\mathsf{old}} = \Sigma^j_{\mathsf{old}}$ and $\Theta^i_{\mathsf{old}} = \Theta^k_{\mathsf{old}}$. [5] For 1k-PMAC_Plus, Σ^i can collide with some previous value of Σ (i.e. $\Sigma^i = \Sigma^j$) as well as some internal input value of the hash computations (i.e. $\Sigma^i = X^j_\alpha$, where $j \leq i$). Similarly, Θ^i can collide with either some Θ^j $(j < i)$ or with some X^j_α. Thus, it gives rise to four different cases i.e.:

- $\Sigma^i = X^j_\alpha, \Theta^i = X^k_\alpha$
- $\Sigma^i = X^j_\alpha, \Theta^i = \Theta^k$
- $\Sigma^i = \Sigma^j, \Theta^i = X^k_\alpha$
- $\Sigma^i = \Sigma^j, \Theta^i = \Theta^k$

[4] One could see this definition as an extended version of the definition of cover-free tuple, defined in [ZWSW12], which says that $(\Sigma^{[q]}, \Theta^{[q]})$ is said to be a covered tuple if $\exists i \in [q]$ such that Σ^i is non-fresh in $\Sigma^{[q]}$ and Θ^i is non-fresh in $\Theta^{[q]}$. If no such i exists then the tuple is said to be cover-free tuple.

[5] This event was named Coll* in [Yas11].

5.1.2 Σ^i is non-fresh and Θ^i is fresh

In this case the output is uniformly random unless the sampled output of Θ^i collides with some range value. This leads to the following three different cases:

- If $\Sigma^i = \Sigma^j$ for some $j \neq i$ (but not equals to some internal values X^j_α), and the sampled output of Θ^i collides with some range value then we call this bad event RCOLL as defined in Fig. 5.2. [6]

- If Σ^i collides with some internal input values X^j_α, then if $Y^j_\alpha \oplus T^i$, which is to be assigned to the output of Θ^i, collides with some internal output values Y^k_β then the bad event happens and we call this bad event PCF1 (defined in Fig. 5.2).

- If Σ^i collides with some internal input values X^j_α then if $Y^j_\alpha \oplus T^i$, which is to be assigned to the output of Θ^i becomes equal to $Y^l_\beta \oplus T^k$ where either Σ^k or Θ^k collides with some internal input value X^l_β then the bad event happens and we call this bad event PCF2 (defined in Fig. 5.2).

5.1.3 Θ^i is non-fresh and Σ^i is fresh

This case is similar to Case 5.1.2, only the role of Σ^i and Θ^i is interchanged.

5.1.4 Both Σ^i and Θ^i are fresh

This case is similar to Case (d) of [Yas11] with a subtle difference. For PMAC_Plus, Yasuda used Lucks's result [Luc00] on the sum of two independent permutations as the construction uses three independent keys. As we move to the single key setting, we require a more general theorem on the conditional distribution of the sum of two identical random permutation (see Theorem 2). [7] We need the conditional distribution of the sum of two identical random permutation as some inputs-outputs of the random permutation Π have already been fixed earlier due to the internal hash computation.

5.2 Proof of the Theorem 3

Using a standard argument, we analyze the security of the construction 1k-PMAC_Plus[Π] (denoted by 1k-PP) based on random permutation Π, instead of the keyed block cipher. This conversion will add a term $\mathbf{Adv}^{\mathrm{prp}}_E(\sigma', t')$ in the advantage, where $\sigma' = \sigma + 2 + 2q$ and $t' = t + O(\sigma + 2 + 2q)$. Therefore, we show that,

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{1k\text{-}PP}}(q, \sigma) \leq \frac{21\sigma}{N} + \frac{224q\sigma^2}{N^2}.$$

Now as we are bounding the prf advantage of 1k-PP information theoretically, we do not consider the time parameter of a distinguisher and hence wlog we consider the *deterministic and unbounded distinguisher.* In the remaining of the section, we prove Theorem 3. In Sect. 5.2.1, we discuss the power of a distinguisher and the description of ideal oracle. We define the set of bad transcripts and bound the probability of it in Sect. 5.2.2. We analyse good transcripts in Sect. 5.2.3.

[6] Yasuda [Yas11] named this event as UpLow*.

[7] For one-key construction, there is no way to have unconditional distribution of sum of random permutation. Thus, we cannot apply directly the result by Bellare et al. [BI99] or by Patarin [Pat08b]. In one key construction, we can obtain the hash value only if we fix the computation of H^π which requires the condition on π. So the sum of permutation for random permutation with some loss of entropy is essential for one key.

5.2.1 Initial Set-up

We fix a deterministic non-repeating query making distinguisher $\mathcal{A}$ that interacts with either (1) the real oracle $\mathsf{1k\text{-}PP}$ for a random permutation Π or (2) the ideal oracle \$, making at most q queries with an aggregate of total σ many message blocks.

Description of Ideal Oracle. An ideal oracle consists of the following two phases: (a) In the online phase, for each query M^i, the oracle samples the response T^i from $\mathbb{B}$ uniformly at random and returns it to $\mathcal{A}$. (b) In the offline phase, in which after $\mathcal{A}$ makes all the queries responses, it first samples Δ_0, Δ_1 from $\mathbb{B}$ in without replacement manner. Then it samples the internal hash value for all the queries in without replacement manner from $\mathbb{B}$. During this sampling stage, if some specfic event occurs (as shown inside the box in Fig. 5.1), then it aborts the sampling process. More formally, ideal oracle \$ works as shown in Fig. 5.1.

Description of Attack Transcript. Let $\tau := (M^{[q]}, T^{[q]})$ be the list of queries and responses of $\mathcal{A}$ that constitutes the query response transcript of $\mathcal{A}$. For convenience, we slightly modify the experiment where we reveal to the distinguisher $\mathcal{A}$ (after it made all its queries and obtains corresponding responses but before it output its decision) the transcript of internal computations $(X, Y, \Sigma^{[q]}_{\mathsf{out}}, \Theta^{[q]}_{\mathsf{out}})$ (this is obviously wlog since the distinguisher can ignore this additional piece of information). If $\mathcal{A}$ interacts with the real oracle, we have $\Sigma^i_{\mathsf{out}} = \Pi(\Sigma^i)$ and $\Theta^i_{\mathsf{out}} = \Pi(\Theta^i)$ for all $i \in [q]$ and (X, Y) is permutation compatible, denoted as $X \xrightarrow{\Pi} Y$. All in all, the transcript of the distinguisher $\mathcal{A}$ is $\tau := (M^{[q]}, T^{[q]}, X, Y, \Sigma^{[q]}_{\mathsf{out}}, \Theta^{[q]}_{\mathsf{out}})$. Note that for such a transcript τ, in the real world we must have

$$\begin{aligned}&\Sigma^i_{\mathsf{out}} \oplus \Theta^i_{\mathsf{out}} = T^i, \ \forall i \in [q], \quad \text{and} \\ &(X,\ \Sigma^{[q]},\ \Theta^{[q]}) \xmapsto{\Pi} (Y,\ \Sigma^{[q]}_{\mathsf{out}},\ \Theta^{[q]}_{\mathsf{out}}).\end{aligned}$$

Here we use the fact that Σ and Θ can be computed from Y. A transcript τ is said to be *attainable* (with respect to distinguisher $\mathcal{A}$) if the probability to obtain this transcript in the ideal world is non zero. Recall that, $\mathcal{V}$ denotes the set of all attainable transcripts and X_{re} and X_{id} denotes the probability distribution of transcript τ induced by the real world and ideal world respectively.

5.2.2 Definition and Probability of Bad Transcripts

In this section, we define and bound the set of bad transcripts in the ideal world. We start by defining the set of bad transcripts.

Definition 2. We say that an attainable transcript $\tau = (M^{[q]}, T^{[q]}, X, Y, \Sigma^{[q]}_{\mathsf{out}}, \Theta^{[q]}_{\mathsf{out}})$ is bad if either of the following bad flags are set to 1: ZeroT, $\mathsf{ZeroOneX}$, ECF, $\mathsf{PCF1}$, $\mathsf{PCF2}$, RCOLL (as defined in Fig. 5.1). With abuse of notation we use the name of the bad flag to denote its corresponding bad event.

Let $\mathcal{V}_b$ be the set of all bad transcripts and $\mathcal{V}_g := \mathcal{V} \setminus \mathcal{V}_b$ be the set of all good transcripts. Now, we define the following event:

$$\mathsf{E\text{-}Bad1} := \mathsf{ZeroT} \ \vee \ \mathsf{ZeroOneX}, \qquad \mathsf{E\text{-}Bad2} := \mathsf{ECF} \ \vee \ \mathsf{PCF1} \ \vee \ \mathsf{PCF2} \ \vee \ \mathsf{RCOLL},$$
$$\mathsf{E\text{-}Bad} \ := \ \mathsf{E\text{-}Bad1} \ \vee \ \mathsf{E\text{-}Bad2}. \tag{4}$$

Now, we bound the probability of realizing the bad transcripts in the ideal world. In specific, to bound the probability of bad transcripts in ideal world, it suffices to bound the probability of the event $\mathsf{E\text{-}Bad}$ (due to Definition 2 and Eqn. (4)) in the following lemma.

```
ONLINE PHASE OF IDEAL ORACLE
$\forall i \in [q]$ : On $i$-th query $M^i$, return $T^i \xleftarrow{\$} \mathbb{B}$;
\\ bad event in online phase
    if $\exists i : T^i = \mathbf{0}$ then $\boxed{\mathsf{ZeroT} \leftarrow 1,}$ $\perp$;
```

```
OFFLINE PHASE OF IDEAL ORACLE, INITIALIZE $\mathcal{L}_{\text{set}} = \mathcal{L}_1 = \mathcal{L}_2 = \emptyset$
    $\mathcal{L}_1(\mathbf{0}) \leftarrow \Delta_0 \xleftarrow{\$} \mathbb{B}$; $\mathcal{L}_1(\mathbf{1}) \leftarrow \Delta_1 \xleftarrow{\$} \mathbb{B} \setminus \{\Delta_0\}$;
    $\forall i \in [q]$ : $(\Sigma^i, \Theta^i) \leftarrow$ $\mathsf{Internal}^{\mathcal{L}_1}(M^i)$
        $\forall j \in [l_i]$ : $X^i_j = 2^j \Delta_0 \oplus 2^{2j} \Delta_1 \oplus M^i_j$;
            if $\mathcal{L}_1(X^i_j) = \top$
                then $\mathcal{L}_1(X^i_j) \leftarrow Y^i_j \xleftarrow{\$} \overline{Ran(\mathcal{L}_1)}$;
            else $Y^i_j \leftarrow \mathcal{L}_1(X^i_j)$;
        $\Sigma^i := \mathsf{fix0}(Y^i_1 \oplus \cdots \oplus Y^i_{l_i})$;
        $\Theta^i := \mathsf{fix1}(2Y^i_1 \oplus 2^2 Y^i_2 \cdots \oplus 2^{l_i} Y^i_{l_i})$;
        return $(\Sigma^i,\ \Theta^i)$;
    if $X^i_j \in \{\mathbf{0}, \mathbf{1}\}$ then $\boxed{\mathsf{ZeroOneX} \leftarrow 1,}$ $\perp$; \\ bad event for $\Delta_0$ and $\Delta_1$ sampling
    if $(\Sigma^{[q]}, \Theta^{[q]})$ is not e.c.f. tuple then $\boxed{\mathsf{ECF} \leftarrow 1,}$ $\perp$; \\ bad event for $\Delta Y$ sampling
    $\forall i \in [q]$ : if $\Sigma^i_{\mathsf{out}} := \mathcal{L}_1(\Sigma^i) \neq \top$ then Case A;
                 if $\Theta^i_{\mathsf{out}} := \mathcal{L}_1(\Theta^i) \neq \top$ then Case B;
    $\mathcal{F}_\Sigma \leftarrow \{i \in [q] : \exists j \neq i,\ \Sigma^i = \Sigma^j\}$; $\mathcal{F}_\Theta \leftarrow \{i \in [q] : \exists j \neq i,\ \Theta^i = \Theta^j\}$;
    $\mathcal{L}_2 = \mathcal{L}_1 \cup \mathcal{L}_{\text{set}}$;    \\ merge two lists. $\mathcal{L}_{\text{set}}$ appeared in Case A, B
    $\forall i \in [q]$ : if $i \in \mathcal{F}_\Sigma$ then Case C;
                 if $i \in \mathcal{F}_\Theta$ then Case D;
    $\mathcal{F} = \{i \in [q] : \mathcal{L}_2(\Sigma^i) = \top = \mathcal{L}_2(\Theta^i)\}$; $f = |\mathcal{F}|$;
    $(\Sigma^i_{\mathsf{out}}, \Theta^i_{\mathsf{out}})_{i \in \mathcal{F}} \xleftarrow{\$} \mathcal{S} := \{(a^i, b^i)_i \in \overline{Ran(\mathcal{L}_2)}^{(2f)} : a^i \oplus b^i = T^i\}$;
    return $(X, Y, \Sigma^{[q]}_{\mathsf{out}}, \Theta^{[q]}_{\mathsf{out}})$;
```

Figure 5.1: Ideal oracle \$: Boxed statements denote bad events. Whenever a bad event is set to 1, the oracle immediately aborts (denoted as $\perp$) and returns the remaining values of the transcript in any arbitrary manner. So, if we proceed further we can surely assume that the event $\perp$ (and so any bad event so far) does not hold. We write $\top$ when the value of a variable is not defined. Shaded box is used to represent the uniform sampling in ideal oracle.

Case A

```
$\Theta^i_{\mathsf{out}} := \Sigma^i_{\mathsf{out}} \oplus T^i$;
if $\Theta^i_{\mathsf{out}} \in Ran(\mathcal{L}_{\text{set}})$ : $\boxed{\mathsf{PCF2} \leftarrow 1,}$ $\bot$;
if $\Theta^i_{\mathsf{out}} \in Ran(\mathcal{L}_1)$ : $\boxed{\mathsf{PCF1} \leftarrow 1,}$ $\bot$;
$\mathcal{L}_{\text{set}}(\Theta^i) = \Theta^i_{\mathsf{out}}$;
```

Case B

```
$\Sigma^i_{\mathsf{out}} := \Theta^i_{\mathsf{out}} \oplus T^i$;
if $\Sigma^i_{\mathsf{out}} \in Ran(\mathcal{L}_{\text{set}})$ : $\boxed{\mathsf{PCF2} \leftarrow 1,}$ $\bot$;
if $\Sigma^i_{\mathsf{out}} \in Ran(\mathcal{L}_1)$ : $\boxed{\mathsf{PCF1} \leftarrow 1,}$ $\bot$;
$\mathcal{L}_{\text{set}}(\Sigma^i) = \Sigma^i_{\mathsf{out}}$;
```

Case C

```
if $\mathcal{L}_2(\Sigma^i) = \top$ : $\boxed{\mathcal{L}_2(\Sigma^i)} \leftarrow_{\$} \overline{Ran(\mathcal{L}_2)}$;
$\Sigma^i_{\mathsf{out}} := \mathcal{L}_2(\Sigma^i)$;
if $\Theta^i_{\mathsf{out}} := \mathcal{L}_2(\Sigma^i) \oplus T^i \in Ran(\mathcal{L}_2)$ :
    $\boxed{\mathsf{RCOLL} \leftarrow 1,}$ $\bot$;
set $\mathcal{L}_2(\Theta^i) = \Theta^i_{\mathsf{out}}$;
```

Case D

```
if $\mathcal{L}_2(\Theta^i) = \top$ : $\boxed{\mathcal{L}_2(\Theta^i)} \leftarrow_{\$} \overline{Ran(\mathcal{L}_2)}$;
$\Theta^i_{\mathsf{out}} := \mathcal{L}_2(\Theta^i)$;
if $\Sigma^i_{\mathsf{out}} := \mathcal{L}_2(\Theta^i) \oplus T^i \in Ran(\mathcal{L}_2)$ :
    $\boxed{\mathsf{RCOLL} \leftarrow 1,}$ $\bot$;
set $\mathcal{L}_2(\Sigma^i) = \Sigma^i_{\mathsf{out}}$;
```

Figure 5.2: Continutation of ideal oracle \$: PCF1 and PCF2 is defined in Case A and Case B. RCOLL is defined in Case C and Case D. We denote the bad event defined in Case C and Case D by RCOLL_1 and RCOLL_2 respectively. $\bot$ and $\top$ denotes the abort symbol and an undefined variable resp. Shaded box represents uniform sampling in ideal oracle.

Lemma 1. *Let X_{id} and $\mathcal{V}_b$ be defined as above then,*

$$\Pr[X_{\text{id}} \in \mathcal{V}_b] \leq \epsilon_{\text{bad}} = \frac{206q\sigma^2}{N^2} + \frac{21\sigma}{N}.$$

Proof. As discussed before, bounding the probability of bad transcripts in the ideal world is equivalent to bounding the probability of the event E-Bad holds in the ideal world. To bound Pr[E-Bad], we bound the probability of the following events:

Bounding E-Bad1.

- For a fixed $i \in [q]$, it is easy to see that $\Pr[T^i = \mathbf{0}] = \frac{1}{N}$ since all T^i's are sampled uniformly at random from $\mathbb{B}$, as defined in Fig. 5.1. Therefore, by varying over all possible choices of i, we obtain $\Pr[\mathsf{ZeroT}] \leq \frac{q}{N}$.
- For fixed $i \in [q], j \in [l_i]$ and $b \in \{\mathbf{0}, \mathbf{1}\}$, $X^i_j = b \Leftrightarrow 2^j \Delta_0 \oplus 2^{2j} \Delta_1 = M^i_j \oplus b$. Therefore, using Corollary 1, we have $\Pr[X^i_j = \mathbf{0}] = \Pr[X^i_j = \mathbf{1}] \leq \frac{1}{N-1}$. Varying over all possible choices of i, j and b, we obtain $\Pr[\mathsf{ZeroOneX}] \leq \frac{2\sigma}{N-1} \leq \frac{3\sigma}{N}$.

Combining the above two, we have $\Pr[\mathsf{E\text{-}Bad1}] \leq \frac{4\sigma}{N}$ ($\because q \leq \sigma$).

Bounding $\mathsf{E\text{-}Bad2} \wedge \overline{\mathsf{E\text{-}Bad1}}$.

- We handle this case in the following lemma, proof of which is postponed to Sect. 7.

  Lemma 2. $\Pr[\mathsf{E\text{-}Bad2} \wedge \overline{\mathsf{E\text{-}Bad1}}] \leq \frac{206q\sigma^2}{N^2} + \frac{17\sigma}{N}$.

The result follows as we sum up the above two bounds.

5.2.3 Analysis of Good Transcripts

Having defined and bounded the probability of realizing bad transcript in the ideal world, it remains to lower bound the ratio of the real and the ideal interpolation probability for a good transcript. For this, let us first understand what a good transcript in ideal oracle means. Note that for each $i \in \mathcal{F}$ (see the definition in line 9 of Fig. 5.1) both Σ^i_{out} and Θ^i_{out} are fresh. As ECF is not 1, for every $i \notin \mathcal{F}$, exactly one of Σ^i_{out} or Θ^i_{out} is fresh. Thus, we have exactly $q - f$ non-fresh blocks and remaining $q + f$ fresh blocks, where $f = |\mathcal{F}|$. We identify two sets $\mathcal{F}^{'}_{\Sigma}$ and $\mathcal{F}^{'}_{\Theta}$ that contain all indices i such that Σ^i collides with some internal input of hash computation or Θ^i collides with some internal input of hash computations respectively. Now, we define an equivalence relation $\sim_\Sigma$ on $\mathcal{F}_\Sigma := [q] \setminus \mathcal{F}^{'}_{\Sigma} \cup \mathcal{F}$ (which is defined in line 6 of Fig. 5.1) as $i \sim_\Sigma j$ if $\Sigma^i = \Sigma^j$. Similarly, we define an equivalence relation $\sim_\Theta$ on $\mathcal{F}_\Theta := [q] \setminus \mathcal{F}^{'}_{\Theta} \cup \mathcal{F}$ as $i \sim_\Theta j$ if $\Theta^i = \Theta^j$. Here we would like to point out that we cannot have $\Sigma^i = \Theta^j$ as we have separated the range of collisions by applying fix_0, fix_1 functions.

Clearly, $\sim_\Sigma$ and $\sim_\Theta$ are equivalence relations on $\mathcal{F}_\Sigma$ and $\mathcal{F}_\Theta$ respectively and hence we can partition the set $\mathcal{F}_\Sigma$ as $C_1 \sqcup \cdots \sqcup C_r$ where each C_j is a part and the set $\mathcal{F}_\Theta$ as $C'_1 \sqcup \cdots \sqcup C'_{r'}$ where C'_j is a part. We call the equivalence class C_j as Σ-class and C'_j as Θ-class. Note that all parts contain at least two elements. Let $c_j = \min C_j$ be the minimum value of partition C_j and so is $c'_j = \min C'_j$. So, when $i = c_j$ or $c'_{j'}$ for some $j \in [r]$ or $j' \in [r']$, we sample the output $\mathcal{L}_2(\cdot)$ (see the definition in line 1 of Case C or Case D respectively), which determines the outputs for all the elements in the corresponding equivalent class C_j or C'_j respectively.

Due to the definition of Σ_{out} and Θ_{out} and the good transcript, we have the following result.

Claim 2. *For a good transcript τ, the following $2q$ tuples of input and output of a random permutation Π, namely,*

$$\begin{aligned} \mathbf{I} &:= (\Sigma^1, \Sigma^2, \ldots, \Sigma^q, \Theta^1, \Theta^2, \ldots, \Theta^q) \\ \mathbf{O} &:= (\Sigma^1_{\mathsf{out}}, \Sigma^2_{\mathsf{out}}, \ldots, \Sigma^q_{\mathsf{out}}, \Theta^1_{\mathsf{out}}, \Theta^2_{\mathsf{out}}, \ldots, \Theta^q_{\mathsf{out}}) \end{aligned}$$

are permutation compatible.

This is true since no range collision occurs for two different inputs as the bad flag events are not set to 1. This observation will help us to bound the real interpolation probability for a good transcript.

Lemma 3. *Let $\tau = (M^{[q]}, T^{[q]}, X, Y, \Sigma^{[q]}_{\mathsf{out}}, \Theta^{[q]}_{\mathsf{out}})$ be a good transcript. Then,*

$$\frac{\Pr[X_{\mathrm{re}} = \tau]}{\Pr[X_{\mathrm{id}} = \tau]} \geq 1 - \frac{18q\sigma^2}{N^2}.$$

Proof. We first note that the tuple $(\Sigma^{[q]}_{\mathsf{out}}, \Theta^{[q]}_{\mathsf{out}})$ is an extended cover-free tuple as τ is a good transcript. Moreover, in Fig. 5.1, we have performed two phases of lazy sampling. In the first phase, we sample the internal outputs of the hash computation through list $\mathcal{L}_1$ (see line 1 of $\mathsf{Internal}$ subroutine in Fig. 5.1). In the next phase, we sample the outputs of Σ^i or Θ^i (as described in line 1 of Case C or D respectively) where $i = c_j$ or $c'_{j'}$ for some $j \in [r]$ or $j' \in [r']$ respectively through list $\mathcal{L}_2$. Let us assume that, the size of the list $\mathcal{L}_1$ is η. We consider the set $\mathcal{F}$ of all free indices as defined in line 9 with $f = |\mathcal{F}|$ and a set $\mathcal{S}$ in line 10 of Fig. 5.1 respectively. We also define a set $\mathcal{I} := \mathcal{F}_\Sigma \cup \mathcal{F}_\Theta \cup \mathcal{F}$. With this notation, we can compute the ideal interpolation probability $p_{\mathrm{id}} := \Pr[X_{\mathrm{id}} = \tau]$ as follows.

$$\begin{aligned} p_{\mathrm{id}} &= \Pr\Big[T^{[q]} = t^{[q]} \wedge \mathcal{L}_1(x^i_j) = y^i_j \wedge \mathcal{L}_2(\Sigma^{i'}) = \Sigma^{i'}_{\mathsf{out}} \wedge \mathcal{L}_2(\Theta^{i'}) = \Theta^{i'}_{\mathsf{out}},\ \forall i' \in \mathcal{I}\Big] \\ &= \frac{1}{N^q} \times \Pr[\underbrace{\mathcal{L}_1(x^i_j) = y^i_j}_{\mathsf{E1}} \wedge \underbrace{\mathcal{L}_2(\Sigma^{i'}) = \Sigma^{i'}_{\mathsf{out}}}_{\mathsf{E2}} \wedge \underbrace{\mathcal{L}_2(\Theta^{i'}) = \Theta^{i'}_{\mathsf{out}}}_{\mathsf{E3}},\ \forall i' \in \mathcal{I}]. \qquad (5) \end{aligned}$$

The first equality follows from the fact that distribution of T^i's are indepedent of the lazy sampling that we carry out in the offline phase of the game. Now, consider the following observations:

- $\Pr\left[\mathcal{L}_1(x^i_j) = y^i_j\right] = \frac{1}{\mathbf{P}(N,\ \eta)}$ as $|\mathcal{L}_1| = \eta$.
- The conditional probability

$$\Pr[\underbrace{\mathsf{E2} \wedge \mathsf{E3},\ \forall i' \in \mathcal{I} \setminus \mathcal{F}}_{\mathsf{E4}} \mid \mathsf{E1}] = \frac{1}{\mathbf{P}(N - (2f + \eta),\ r + r')},$$

as we need to sample the output for a single element from each equivalent class and there are all total $r + r'$ equivalence classes (combining the Σ-class and Θ-class).

- Finally for all free indices i, we sample the output from $\mathcal{S}$. Therefore,

$$\begin{aligned}\Pr[\mathsf{E2} \wedge \mathsf{E3},\ \forall i' \in \mathcal{F} \mid \mathsf{E1} \wedge \mathsf{E4}] \leq \frac{1}{|\mathcal{S}|} &\leq \frac{\mathbf{P}(N - \eta,\ 2f)}{N^f \times (1 - \frac{4f\eta^2 + 8f^2\eta + 6f^3}{N^2})} \\ &\leq \frac{\mathbf{P}(N - \eta,\ 2f)}{N^f \times (1 - \frac{18q\sigma^2}{N^2})}.\end{aligned}$$

This follows from the lower bound of $|\mathcal{S}|$ from Theorem 2 with the assumption $\eta + 2f \leq \frac{N}{2}$, $\eta \leq \sigma$ and $f \leq q \leq \sigma$.

Therefore, we have

$$p_{\text{id}} \leq \frac{1}{N^q} \times \frac{1}{\mathbf{P}(N,\ \eta)} \times \frac{1}{\mathbf{P}(N - (2f + \eta),\ r + r')} \times \frac{\mathbf{P}(N - \eta,\ 2f)}{N^f \times (1 - \frac{18q\sigma^2}{N^2})}. \tag{6}$$

Now, we compute the real interpolation probability for a good transcript τ. By virtue of Claim 2, we know that, $(\Sigma^{[q]}, \Theta^{[q]})$ is permutation compatible with $(\Sigma^{[q]}_{\mathsf{out}},\ \Theta^{[q]}_{\mathsf{out}})$. Note that the number of distinct elements in $\Sigma^{[q]} \| \Theta^{[q]}$ is exactly $q + f + r + r'$. Hence,

$$p_{\text{re}} := \Pr[X_{\text{re}} = \tau] \quad = \quad \frac{1}{\mathbf{P}(N,\ \eta)} \times \frac{1}{\mathbf{P}(N - \eta,\ q + f + r + r')}. \tag{7}$$

Therefore,

$$\frac{\Pr[X_{\text{re}} = \tau]}{\Pr[X_{\text{id}} = \tau]} \geq \frac{N^q \times N^f \times (1 - \frac{18q\sigma^2}{N^2}) \times \mathbf{P}(N - (2f + \eta),\ r + r')}{\mathbf{P}(N - \eta,\ 2f) \times \mathbf{P}(N - \eta,\ q + f + r + r')} \geq (1 - \frac{18q\sigma^2}{N^2}).$$

Applying Theorem 1 with $\epsilon_{\text{ratio}} = \frac{18q\sigma^2}{N^2}$ and $\epsilon_{\text{bad}} = \frac{21\sigma}{N} + \frac{206q\sigma^2}{N^2}$, the result of Theorem 3 follows.

6 Bounding Internal Bad Events for Proving Lemma 2

In the last section, we have proved that 1k-PMAC_Plus is indistinguishable from random function upto close to $2^{2n/3}$ blocks if the underlying block cipher is assumed to be a secure PRP, with keeping Lemma 2 unproven. Thus, it only remains to prove Lemma 2, which we will do in Sect. 7. Before that, in this section, we define and bound some additional internal bad events, which are different from the list of bad events (i.e. E-Bad) already identified in Sect. 5.2.2. As we will see later, these additional internal bad events will help us in proving Lemma 2.

Consider a tuple of q messages $M^1, \ldots, M^q$ and let l_i denote the number of message blocks of message M^i (we assume that all the messages are of size multiple of n). Now, we fix two distinct indices $i, j \in [q]$ and we define a set $\mathsf{NEQ}_{i,j} := \{\alpha \in [\min\{l_i, l_j\}] : M^i_\alpha \neq M^j_\alpha\} \cup \{\alpha : l_j + 1 \leq \alpha \leq l_i\}$. In other words, the set $\mathsf{NEQ}_{i,j}$ contains all the positions, where the message blocks of i-th and j-th message are not equal. Having defined the set, we define the internal bad events in Fig. 6.1.

List of Internal Bad Events

1 : $\mathsf{3CollX} :=$ Fix $i \neq j \in [q]$. $\exists i_1, i_2, i_3 \in \{i, j\} : \alpha \in [l_{i_1}], \beta \in [l_{i_2}]$,
$\gamma = \min \mathsf{NEQ}_{i,j}$ where $\alpha \neq \beta \neq \gamma : X_\alpha^{i_1} = X_\beta^{i_2} = X_\gamma^{i_3}$ (see Remark 1).

2 : $\mathsf{ZeroY} := \exists i \in [q], j \in [l_i]$ such that $Y_j^i = \mathbf{0}$.

3 : $\mathsf{Bad}_1 := \exists i, j, k \in [q], \alpha \in [l_j], \beta \in [l_k] (\neq \alpha)$ such that
$\Sigma^i =_1 X_\alpha^j$ and $X_\alpha^j = X_\beta^k$.

4 : $\mathsf{Bad}_2 := \exists i, j, k \in [q], \alpha \in [l_j], \beta \in [l_k] (\neq \alpha)$ such that
$\Theta^i =_1 X_\alpha^j$ and $X_\alpha^j = X_\beta^k$.

5 : $\mathsf{Bad}_3 := \exists i, j, k \in [q], \alpha \in [l_j], \beta \in [l_k] (\neq \alpha)$ such that
$\Sigma^i =_1 X_\alpha^j$ and $X_\gamma^i = X_\beta^k$, where $\gamma \in \min \mathsf{NEQ}_{i,j}$.

6 : $\mathsf{Bad}_4 := \exists i, j, k \in [q], \alpha \in [l_j], \beta \in [l_k] (\neq \alpha)$ such that
$\Theta^i =_1 X_\alpha^j \oplus b$ and $X_\gamma^i = X_\beta^k$, where $\gamma \in \min \mathsf{NEQ}_{i,j}$.

7 : $\mathsf{Bad}_5 := \exists i \in [q], \alpha \in [l_i - 1] : X_{l_i}^i = X_\alpha^i$.

Figure 6.1: List of Internal Bad Events.

Remark 1. We would like to emphasize that our definition of the 3CollX event (see Fig. 6.1) is substantially different from that of Yasuda's [Yas11]. Yasuda in [Yas11] considered three collisions between three messages and hence obtained the bound $\frac{q^3\ell^3}{N^2}$. But we observe that, it is enough to consider three collisions between a fixed pair of messages and a fixed choice of the message block index. Moreover, according to our definition of 3CollX, choice of γ is unique after the pair of messages are fixed. Hence, we become able to reduce the dependency of length in the security bound from cubic to quadratic.

Having defined all the internal bad events, we define the event

$$\mathsf{I\text{-}Bad} := \mathsf{3CollX} \ \vee \ \mathsf{ZeroY} \ \vee_{a=1}^{5} \mathsf{Bad}_a.$$

Also, recall that we have defined

$$\mathsf{E\text{-}Bad1} := \mathsf{ZeroT} \ \vee \ \mathsf{ZeroOneX}.$$

Now, we have

$$\Pr[\mathsf{I\text{-}Bad} \wedge \overline{\mathsf{E\text{-}Bad1}}] \leq \Pr[\mathsf{3CollX} \ |\overline{\mathsf{E\text{-}Bad1}}] + \Pr[\mathsf{ZeroY} \ |\overline{\mathsf{E\text{-}Bad1}}] + \sum_{a=1}^{5} \Pr[\mathsf{Bad}_a \ |\overline{\mathsf{E\text{-}Bad1}}]. \quad (8)$$

Now, we bound all the internal bad events that we have identified in Fig. 6.1 conditioned on $\overline{\mathsf{E\text{-}Bad1}}$, separately. Then using Eqn. (8) we obtain the bound of the probability of $\mathsf{I\text{-}Bad} \wedge \overline{\mathsf{E\text{-}Bad1}}$ as shown in the following Lemma.

Lemma 4. $\Pr[\mathsf{I\text{-}Bad} \wedge \overline{\mathsf{E\text{-}Bad1}}] \leq \frac{3\sigma}{N} + \frac{\sigma^2}{N^2} + \frac{8q\sigma^2}{N^2}$.

Proof. We bound the probability of all the internal bad events separately as follows:

Bounding 3CollX | $\overline{\textbf{E-Bad1}}$. Fix $i \neq j \in [q]$. For any fixed $i_1, i_2, i_3 \in \{i, j\}$ and $\alpha \in [l_{i_1}], \beta \in [l_{i_2}]$, the set of equations $X_\alpha^{i_1} = X_\beta^{i_2}, X_\beta^{i_2} = X_\gamma^{i_3}$ (i.e. $M_\alpha^{i_1} \oplus M_\beta^{i_2} = (2^\alpha \oplus 2^\beta)\Delta_0 \oplus (2^{2\alpha} \oplus 2^{2\beta})\Delta_1$ and $M_\gamma^{i_3} \oplus M_\beta^{i_2} = (2^\gamma \oplus 2^\beta)\Delta_0 \oplus (2^{2\gamma} \oplus 2^{2\beta})\Delta_1$) has always rank 2 as α, β and γ are distinct. Now, using part (b) of Corollary 1, we have

$$\Pr[X_\alpha^{i_1} = X_\beta^{i_2} \wedge X_\beta^{i_2} = X_\gamma^{i_3} \mid \overline{\mathsf{E\text{-}Bad1}}] \leq \frac{1}{N(N-1)}.$$

Summing over all possible choices of i, j and all possible choices of $\alpha \in [l_{i_1}], \beta \in [l_{i_2}]$ we obtain the bound to be $\frac{\sigma(\sigma-1)}{N(N-1)} \leq \frac{\sigma^2}{N^2}$.

Bounding $\textsf{ZeroY} \mid \overline{\textsf{E-Bad1}}$. For a fixed $i \in [q]$ and $\alpha \in [l_i]$,

$$\Pr[Y^i_\alpha = \mathbf{0} \mid \overline{\textsf{E-Bad1}}] = \Pr[\Pi(X^i_\alpha) = \mathbf{0} \mid \overline{\textsf{E-Bad1}}] \leq \frac{1}{N-\sigma}.$$

Varying over all possible choices of i and α we bound this event by $\frac{2\sigma}{N}$, assuming $\sigma \leq \frac{N}{2}$.

Bounding $\textsf{Bad}_5 \mid \overline{\textsf{E-Bad1}}$. Fix $i \in [q]$ and $\alpha \in [l_i - 1]$. As $l_i \neq \alpha$, $2^{l_i} \oplus 2^\alpha \neq \mathbf{0}$. Similarly, $2^{2l_i} \oplus 2^{2\alpha} \neq \mathbf{0}$. Now,

$$\begin{aligned}\Pr[X^i_{l_i} = X^i_\alpha \mid \overline{\textsf{E-Bad1}}] &= \Pr[(2^{l_i} \oplus 2^\alpha)\Delta_0 \oplus (2^{2l_i} \oplus 2^{2\alpha})\Delta_1 = M^i_{l_i} \oplus M^i_\alpha \mid \overline{\textsf{E-Bad1}}] \\ &\leq \frac{1}{N-1}.\end{aligned}$$

The last inequality follows from Corollary 1. By varying over all choices of i and α we obtain the bound to be $\frac{\sigma-1}{N-1} \leq \frac{\sigma}{N}$. Moreover, observe that we require to condition on $\overline{\textsf{E-Bad1}}$, otherwise the event $\textsf{ZeroOneX}$ implies the collision $X^i_{l_i} = X^i_\alpha$ trivially by choosing appropriate messages.

Bounding $\textsf{Bad}_1 \vee \ldots \vee \textsf{Bad}_4 \mid \overline{\textsf{E-Bad1}}$. To bound the event, we first bound the probability of $\textsf{Bad}_1 \mid \overline{\textsf{E-Bad1}}$ as follows:

$$\begin{aligned}\Pr[\textsf{Bad}_1 \mid \overline{\textsf{E-Bad1}}] &= \sum_{i,j,k}\sum_{\alpha,\beta} \Pr[\Sigma^i =_1 X^j_\alpha \wedge X^j_\alpha = X^k_\beta \mid \overline{\textsf{E-Bad1}}] \\ &= \sum_{i,j,k}\sum_{\alpha,\beta} \underbrace{\Pr[\Sigma^i =_1 X^j_\alpha \,|X^j_\alpha = X^k_\beta \wedge \overline{\textsf{E-Bad1}}]}_{(1)} \cdot \underbrace{\Pr[X^j_\alpha = X^k_\beta \mid \overline{\textsf{E-Bad1}}]}_{(2)}.\end{aligned}$$

Now, we make the following claim, proof of which is postponed to Appendix A.

Claim 3. Let M^i, M^j and M^k be three messages. Let $\alpha \in [l_j], \beta \in [l_k]$ and c be a non-zero constant. Then, for any $b, b' \in \{\mathbf{0}, \mathbf{1}\}$, we have

$$\begin{aligned}(a)\,\Pr[\Sigma^i = X^j_\alpha \oplus b \,|cX^j_\alpha = X^k_\beta \oplus b'] &\leq \frac{2}{N}, \\ (b)\,\Pr[\Theta^i = X^j_\alpha \oplus b \,|cX^j_\alpha = X^k_\beta \oplus b'] &\leq \frac{2}{N},\end{aligned}$$

where $\sigma \leq \frac{N}{2}$.

Now, based on the above claim, we consider the following two observations:

- Assuming $\sigma \leq \frac{N}{2}$, (1) can be bounded by $\frac{2}{N}$ (follows from part (a) of Claim 3).
- (2) can by bounded by $\frac{1}{N-1}$, which follows directly from Corollary 1.

Now varying over all choices of i, j, k and α, β, we obtain the bound to be $\frac{2q\sigma(\sigma-1)}{N(N-1)} \leq \frac{2q\sigma^2}{N^2}$.

With similar argument and part (a) of Claim 3, one can show that $\Pr[\textsf{Bad}_3 \mid \overline{\textsf{E-Bad1}}] \leq \frac{2q\sigma^2}{N^2}$ and using part (b) of Claim 3, one can show that $\Pr[\textsf{Bad}_a \mid \overline{\textsf{E-Bad1}}] \leq \frac{2q\sigma^2}{N^2}$ for $a = 2, 4$. The result follows as we put all these bounds in Eqn. (8).

7 Proof of Lemma 2 and Bounding RCOLL, ECF, PCF1, and PCF2

Having defined and bounded all the internal bad events as identified in Sect. 6, we are now ready to prove Lemma 2. We quickly recall the following bad events from Sect. 5.2.2 and Sect. 6

$$\mathsf{E\text{-}Bad1} := \mathsf{ZeroT} \ \vee \ \mathsf{ZeroOneX}, \qquad \mathsf{E\text{-}Bad2} := \mathsf{ECF} \ \vee \ \mathsf{PCF1} \ \vee \ \mathsf{PCF2} \ \vee \ \mathsf{RCOLL},$$
$$\mathsf{I\text{-}Bad} \ := \ \mathsf{3CollX} \ \vee \ \mathsf{ZeroY} \ \vee_{k=1}^{5} \mathsf{Bad}_k.$$

We begin this section with the proof of Lemma 2.

Lemma 2. $\Pr[\mathsf{E\text{-}Bad2} \wedge \overline{\mathsf{E\text{-}Bad1}}] \leq \frac{204q\sigma^2}{N^2} + \frac{17\sigma}{N}$.

Proof. Let us define the following event: $\mathsf{Bad} := \mathsf{E\text{-}Bad1} \vee \mathsf{I\text{-}Bad}$. So, $\overline{\mathsf{Bad}} = \overline{\mathsf{E\text{-}Bad1}} \wedge \overline{\mathsf{I\text{-}Bad}}$. Now, we can write

$$\begin{aligned}
\Pr[\mathsf{E\text{-}Bad2} \wedge \overline{\mathsf{E\text{-}Bad1}}] &\leq \Pr[\mathsf{E\text{-}Bad2} \wedge \overline{\mathsf{E\text{-}Bad1}} \wedge \overline{\mathsf{I\text{-}Bad}}] + \Pr[\mathsf{I\text{-}Bad} \wedge \overline{\mathsf{E\text{-}Bad1}}] \\
&= \Pr[\mathsf{E\text{-}Bad2} \wedge \overline{\mathsf{Bad}}] + \Pr[\mathsf{I\text{-}Bad} \wedge \overline{\mathsf{E\text{-}Bad1}}] \\
&\overset{[1]}{\leq} \Pr[\mathsf{ECF} \wedge \overline{\mathsf{Bad}}] + \Pr[\mathsf{PCF1} \wedge \overline{\mathsf{Bad}}] + \Pr[\mathsf{PCF2} \wedge \overline{\mathsf{Bad}}] \\
&\quad + \Pr[\mathsf{RCOLL} \wedge \overline{\mathsf{Bad}}] + (\frac{3\sigma}{N} + \frac{\sigma^2}{N^2} + \frac{8q\sigma^2}{N^2}) \\
&\overset{[2]}{\leq} \frac{98q\sigma^2}{N^2} + \frac{26q\sigma^2}{N^2} + (\frac{147q^2\sigma^2}{N^3} + \frac{7\sigma}{N}) \\
&\quad + \frac{6\sigma}{N} + (\frac{3\sigma}{N} + \frac{\sigma^2}{N^2} + \frac{8q\sigma^2}{N^2}) \\
&\overset{[3]}{\leq} \frac{206q\sigma^2}{N^2} + \frac{17\sigma}{N},
\end{aligned} \tag{9}$$

where [1] follows from Lemma 4 and [2] follows from Tab. 2. Moreover [3] follows from simple algebraic calculations assuming $q \leq \sigma \leq \frac{N}{2}$.

Table 2: List of the events to be bounded with their corresponding bound.

Event to be Bounded	Bounds of the Event	Reference in the Paper
$\mathsf{RCOLL} \wedge \overline{\mathsf{Bad}}$	$\frac{6\sigma}{N}$	Section 7.1
$\mathsf{ECF} \wedge \overline{\mathsf{Bad}}$	$\frac{98q\sigma^2}{N^2}$	Section 7.2
$\mathsf{PCF1} \wedge \overline{\mathsf{Bad}}$	$\frac{26q\sigma^2}{N^2}$	Section 7.3
$\mathsf{PCF2} \wedge \overline{\mathsf{Bad}}$	$\frac{147q^2\sigma^2}{N^3} + \frac{7\sigma}{N}$	Section 7.4

In the remainder of the section, we bound the probability of the four events (i.e. $\mathsf{RCOLL} \wedge \overline{\mathsf{Bad}}$, $\mathsf{ECF} \wedge \overline{\mathsf{Bad}}$, $\mathsf{PCF1} \wedge \overline{\mathsf{Bad}}$ and $\mathsf{PCF2} \wedge \overline{\mathsf{Bad}}$) as mentioned in Tab. 2. In Sect. 7.1 we establish the bound of the event $\mathsf{RCOLL} \wedge \overline{\mathsf{Bad}}$. Sect. 7.2 is devoted for bounding the event $\mathsf{ECF} \wedge \overline{\mathsf{Bad}}$. $\mathsf{PCF1} \wedge \overline{\mathsf{Bad}}$ and $\mathsf{PCF2} \wedge \overline{\mathsf{Bad}}$ are bounded in Sect. 7.3 and Sect. 7.4 respectively.

7.1 Bounding Joint Probability of RCOLL and $\overline{\mathsf{Bad}}$

In this section, we concentrate on bounding the joint probability of RCOLL and $\overline{\mathsf{Bad}}$. Recall that RCOLL event is triggered from Case C or Case D (refer Fig. 5.2). We separate these two cases in two sub-events, RCOLL_1 and RCOLL_2 respectively, as shown in Tab. 3. Before bounding these two sub-events, we first state an important claim, proof of which can be found in Appendix B.

Table 3: Bound for the joint event (i) RCOLL_1 and $\overline{\mathsf{Bad}}$ and (ii) RCOLL_2 and $\overline{\mathsf{Bad}}$.

Events	Bound
$\mathsf{RCOLL}_1 := (\Sigma^i =_1 \Sigma^j) \wedge (\Theta^i_{\mathsf{out}} \in Ran(\mathcal{L}_2)) \wedge \overline{\mathsf{Bad}}$	$\frac{4\sigma}{N}$
$\mathsf{RCOLL}_2 := (\Theta^i =_1 \Theta^j) \wedge (\Sigma^i_{\mathsf{out}} \in Ran(\mathcal{L}_2)) \wedge \overline{\mathsf{Bad}}$	$\frac{2\sigma}{N}$

Claim 4. *Let M^i and M^j be two distinct messages. If $\sigma \leq \frac{N}{2}$,*

$$(a)\Pr[\Sigma^i =_1 \Sigma^j, \overline{\mathsf{Bad}}] \leq \frac{4(\max\{l_i, l_j\}+1)}{N} \quad \text{and} \quad (b)\Pr[\Theta^i =_1 \Theta^j, \overline{\mathsf{Bad}}] \leq \frac{4}{N}.$$

Now, we bound the joint probability of the event RCOLL_1 and $\overline{\mathsf{Bad}}$ as follows:

$$\begin{aligned}
\Pr[\mathsf{RCOLL}_1 \wedge \overline{\mathsf{Bad}}] &= \sum_{i,j} \Pr[\Sigma^i =_1 \Sigma^j \wedge \Theta^i_{\mathsf{out}} \in Ran(\mathcal{L}_2) \wedge \overline{\mathsf{Bad}}] \\
&\overset{[1]}{=} \sum_{i,j} \Pr[\Sigma^i =_1 \Sigma^j \wedge \overline{\mathsf{Bad}}] \cdot \Pr[\Theta^i_{\mathsf{out}} \in Ran(\mathcal{L}_2)] \\
&\overset{[2]}{\leq} \sum_{i,j} \frac{4(\max\{l_i, l_j\}+1)}{N} \cdot \frac{1}{N-(2q+\eta)} \\
&\leq \frac{4q(\sigma+q)}{N^2} \leq \frac{4\sigma}{N}, \qquad (10)
\end{aligned}$$

where [1] follows from the independence of the two events and [2] follows from Claim 4 and the maximum size of $Ran(\mathcal{L}_2)$ is $2q+\eta$. The last inequality follows from $q \leq \sigma \leq \frac{N}{2}$. With similar arguments one can show,

$$\Pr[\mathsf{RCOLL}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{2\sigma}{N}. \qquad (11)$$

Proof of Eqn. (11) can be found in Appendix D.1.

Now, combining Eqn. (10) and Eqn. (11), we have

$$\Pr[\mathsf{RCOLL} \wedge \overline{\mathsf{Bad}}] \leq \Pr[\mathsf{RCOLL}_1 \wedge \overline{\mathsf{Bad}}] + \Pr[\mathsf{RCOLL}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{6\sigma}{N}.$$

7.2 Bounding Joint Probability of ECF and $\overline{\mathsf{Bad}}$

In this section, we bound the joint probability of ECF and $\overline{\mathsf{Bad}}$. We classify the event ECF into four disjoint events as listed in Tab. 4. To establish the bound for the joint probability of ECF and $\overline{\mathsf{Bad}}$, we separately bound the joint probability of ECF_a and $\overline{\mathsf{Bad}}$ where $a = 1, 2, 3, 4$ and then apply the union bound. In the following analysis, we assume that $\sigma \leq \frac{N}{2}$.

First, we bound the joint probability of ECF_1: $\Sigma^i =_1 \Sigma^j \wedge \Theta^i =_1 X^k_\beta \wedge \overline{\mathsf{Bad}}$. Recall that for fixed $i, j \in [q]$ we defined $\mathsf{NEQ}_{i,j} := \{\alpha \in [\min\{l_i, l_j\}] : M^i_\alpha \neq M^j_\alpha\} \cup \{\alpha : l_j + 1 \leq \alpha \leq l_i\}$. Let $\gamma = \min \mathsf{NEQ}_{i,j}$. Clearly, $\gamma \leq \max\{l_i, l_j\}$ and wlog let us assume that $l_i \leq l_j$. Now, we write the two events (i.e. $\Sigma^i =_1 \Sigma^j$ and $\Theta^i =_1 X^k_\beta$) in terms of Y-variables in the following matrix form:

$$\underbrace{\begin{pmatrix} \mathbf{1} & b & \cdots \\ 2^{l_i-\gamma+1} & X^k_\beta \oplus b' & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_\gamma \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

Table 4: Bound for the joint event (i) $\mathsf{ECF}_1 \wedge \overline{\mathsf{Bad}}$ (ii) $\mathsf{ECF}_2 \wedge \overline{\mathsf{Bad}}$ (iii) $\mathsf{ECF}_3 \wedge \overline{\mathsf{Bad}}$ and (iv) $\mathsf{ECF}_4 \wedge \overline{\mathsf{Bad}}$.

Events	Bound
$\mathsf{ECF}_1 := (\Sigma^i =_1 \Sigma^j) \wedge (\Theta^i =_1 X^k_\beta) \wedge \overline{\mathsf{Bad}}$	$\frac{19q\sigma^2}{N^2}$
$\mathsf{ECF}_2 := (\Sigma^i =_1 X^j_\alpha) \wedge (\Theta^i =_1 \Theta^k) \wedge \overline{\mathsf{Bad}}$	$\frac{19q\sigma^2}{N^2}$
$\mathsf{ECF}_3 := (\Sigma^i =_1 X^j_\alpha) \wedge (\Theta^i =_1 X^k_\beta) \wedge \overline{\mathsf{Bad}}$	$\frac{19q\sigma^2}{N^2}$
$\mathsf{ECF}_4 := (\Sigma^i =_1 \Sigma^j) \wedge (\Theta^i = \Theta^k) \wedge \overline{\mathsf{Bad}}$	$\frac{41q\sigma^2}{N^2}$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Let us define B to be the event $(X^k_\beta = 2^{l_i-\gamma+1} \oplus b')$. It is easy to see that, if $(b = \mathbf{1})$ and B holds, then $rank(A) \geq 1$, otherwise $rank(A) = 2$. Now, we bound the probability of ECF_1 using the above observations as follows:

$$\begin{aligned}\Pr[\mathsf{ECF}_1 \wedge \overline{\mathsf{Bad}}] &\leq \sum_{i,j,k}\sum_{\beta}(\Pr[\Sigma^i = \Sigma^j \wedge \Theta^i =_1 X^k_\beta \;|\overline{\mathsf{Bad}}] \\ &\quad + \Pr[\Sigma^i = \Sigma^j \oplus \mathbf{1} \wedge \Theta^i =_1 X^k_\beta \;|\overline{\mathsf{Bad}} \wedge \overline{\mathsf{B}}]) \\ &\quad + \Pr[\Sigma^i = \Sigma^j \oplus \mathbf{1} \wedge \Theta^i =_1 X^k_\beta \mid \overline{\mathsf{Bad}} \wedge \mathsf{B}] \cdot \Pr[\mathsf{B}] \\ &\overset{[1]}{\leq} \sum_{i,j,k}\sum_{\beta}\left(\frac{2}{(N-\sigma)^2} + \frac{2}{(N-\sigma)^2} + \frac{2}{N}\cdot\frac{1}{N-1}\right) \leq \frac{19q\sigma^2}{N^2}, \end{aligned} \tag{12}$$

where [1] follows from Proposition 1 and the last inequality follows from $q \leq \sigma \leq \frac{N}{2}$. Similar analysis holds for both ECF_2 and ECF_3, and we have

$$\Pr[\mathsf{ECF}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{19q\sigma^2}{N^2}. \tag{13}$$

$$\Pr[\mathsf{ECF}_3 \wedge \overline{\mathsf{Bad}}] \leq \frac{19q\sigma^2}{N^2}. \tag{14}$$

Proof of Eqn. (13) and Eqn. (14) can be found in Appendix D.2 and D.3.

Now, we are left with bounding the joint probability of ECF_4: $\Sigma^i =_1 \Sigma^j \wedge \Theta^i =_1 \Theta^k \wedge \overline{\mathsf{Bad}}$, which requires a different treatment. For that, let CollX_{ijk} denote the event

$$\mathsf{CollX}_{ijk} := X^{i_1}_\alpha = X^{i_2}_\beta,$$

where $i_1, i_2 \in \{i, j, k\}$ and $\alpha \in \{l_{i_1}, \min \mathsf{NEQ}_{i_1,i_2}, \min_2 \mathsf{NEQ}_{i_1,i_2}\}$, $\beta \in [l_{i_2}]$ are distinct. From Corollary 1, it is easy to see that

$$\Pr[\mathsf{CollX}_{ijk}] \leq \frac{3 \cdot \max\{l_i, l_j, l_k\}}{N-1}.$$

Now, we make the following claim, proof of which can be found in Appendix C.

Claim 5. *If* $\overline{\mathsf{CollX}_{\mathbf{ijk}}}$ *occurs, then the system of equations* $\Sigma^i =_1 \Sigma^j$ and $\Theta^i =_1 \Theta^k$ *has rank exactly* 2.

It is easy to see that, if CollX_{ijk} occurs, then the system of equations will have rank at least 1. Based on this claim, we have

$$\begin{aligned}\Pr[\mathsf{ECF}_4 \wedge \overline{\mathsf{Bad}}] &\leq \sum_{i,j,k}\Pr[\Sigma^i =_1 \Sigma^j \wedge \Theta^i =_1 \Theta^k \;|\overline{\mathsf{Bad}} \wedge \overline{\mathsf{CollX}_{ijk}}] \\ &\quad + \Pr[\Sigma^i =_1 \Sigma^j \wedge \Theta^i =_1 \Theta^k \mid \mathsf{CollX}_{ijk} \wedge \overline{\mathsf{Bad}}] \cdot \Pr[\mathsf{CollX}_{ijk}] \\ &\overset{[1]}{\leq} \sum_{i,j,k}\left(\frac{4}{(N-\sigma)^2} + \frac{4}{N-\sigma}\cdot\frac{3\cdot\max\{l_i, l_j, l_k\}}{N-1}\right) \leq \frac{41q\sigma^2}{N^2}, \end{aligned} \tag{15}$$

where [1] follows from Proposition 1 and the last inequality follows from the assumption $q \leq \sigma \leq \frac{N}{2}$ with simple algebraic calculation.
Finally, combining Eqn. (12), Eqn. (13), Eqn. (14) and Eqn. (15), we have

$$\Pr[\mathsf{ECF} \wedge \overline{\mathsf{Bad}}] \leq \sum_{a=1}^{4} \Pr[\mathsf{ECF}_a, \overline{\mathsf{Bad}}] \leq \frac{98q\sigma^2}{N^2}.$$

7.3 Bounding Joint Probability of PCF1 and $\overline{\mathsf{Bad}}$

In this section, we bound the joint probability of $\mathsf{PCF1}$ and $\overline{\mathsf{Bad}}$. As before, we classify $\mathsf{PCF1}$ into two disjoint events and bound them separately. In the following analysis, we assume that $\sigma \leq \frac{N}{2}$.

Table 5: Bound for the joint event (i) $\mathsf{PCF1}_1 \wedge \overline{\mathsf{Bad}}$ and (ii) $\mathsf{PCF1}_2 \wedge \overline{\mathsf{Bad}}$.

Events	Bound
$\mathsf{PCF1}_1 := (\Sigma^i =_1 X^j_\alpha) \wedge (Y^j_\alpha \oplus Y^k_\beta = T^i) \wedge \overline{\mathsf{Bad}}$	$\frac{13q\sigma^2}{N^2}$
$\mathsf{PCF1}_2 := (\Theta^i =_1 X^j_\alpha) \wedge (Y^j_\alpha \oplus Y^k_\beta = t^i) \wedge \overline{\mathsf{Bad}}$	$\frac{13q\sigma^2}{N^2}$

To bound the joint probability of $\mathsf{PCF1}_1$ and $\overline{\mathsf{Bad}}$, we represent the two equations corresponding to the event $\mathsf{PCF1}_1$ (i.e. $\Sigma^i =_1 X^j_\alpha$ and $Y^j_\alpha \oplus Y^k_\beta = T^i$) in terms of Y-variables in the following matrix form:

$$\underbrace{\begin{pmatrix} \mathbf{1} & X^j_\alpha \oplus b & \cdots \\ \mathbf{0/1} & T^i & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_{l_i} \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

where $b \in \{\mathbf{0}, \mathbf{1}\}$. Let us denote the event $X^j_\alpha = T^i \oplus b$ by B. It is easy to see that, if $(Y^i_{l_i} = Y^j_\alpha$ or $Y^i_{l_i} = Y^k_\beta)$ and B holds, then $rank(A) \geq 1$, otherwise $rank(A) = 2$. Hence, we can bound the joint probability as follows:

$$\begin{aligned} \Pr[\mathsf{PCF1}_1 \wedge \overline{\mathsf{Bad}}] &\leq \sum_{i,j,k} \sum_{\alpha,\beta} (\Pr[\mathsf{PCF1}_1 \; | \overline{\mathsf{Bad}} \wedge \overline{\mathsf{B}}] \\ &\quad + \Pr[\mathsf{PCF1}_1 \wedge \overline{\mathsf{Bad}} \mid \mathsf{B}] \cdot \Pr[\mathsf{B}]) \\ &\overset{[1]}{\leq} \sum_{i,j,k} \sum_{\alpha,\beta} \max\left\{ \frac{2}{(N-\sigma)^2}, \left(\frac{2}{(N-\sigma)^2} + \frac{2}{N-\sigma} \cdot \frac{1}{N-1} \right) \right\} \\ &\leq \frac{13q\sigma^2}{N^2}, \end{aligned} \tag{16}$$

where [1] follows from Proposition 1 and the last inequality follows from $q \leq \sigma \leq \frac{N}{2}$ with simple algebraic calculation.
With similar arguments, one can prove that

$$\Pr[\mathsf{PCF1}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{13q\sigma^2}{N^2}. \tag{17}$$

For the sake of completeness, we have provided the proof of Eqn. (17) in Appendix D.4. By combining Eqn. (16) and Eqn. (17), we have

$$\Pr[\mathsf{PCF1} \wedge \overline{\mathsf{Bad}}] \leq \Pr[\mathsf{PCF1}_1 \wedge \overline{\mathsf{Bad}}] + \Pr[\mathsf{PCF1}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{26q\sigma^2}{N^2}.$$

7.4 Bounding Joint Probability of PCF2 and $\overline{\mathsf{Bad}}$

In this section, we bound the joint probability of $\mathsf{PCF2}$ and $\overline{\mathsf{Bad}}$. As before, we classify the event $\mathsf{PCF2}$ into three disjoint events as listed in Tab. 6. To establish the bound for the joint probability of $\mathsf{PCF2}$ and $\overline{\mathsf{Bad}}$, we separately bound the joint probability of $\mathsf{PCF2}_a$ and $\overline{\mathsf{Bad}}$ where $a = 1, 2, 3$ and then apply the union bound. In the following analysis, we assume that $\sigma \leq \frac{N}{2}$.

Table 6: Bound for the joint event (i) $\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}}$ (ii) $\mathsf{PCF2}_2 \wedge \overline{\mathsf{Bad}}$ and (iii) $\mathsf{PCF2}_3 \wedge \overline{\mathsf{Bad}}$.

Events	Bound
$\mathsf{PCF2}_1 := (\Sigma^i =_1 X^j_\alpha) \wedge (\Sigma^k =_1 X^l_\beta) \wedge (Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k) \wedge \overline{\mathsf{Bad}}$	$\frac{49q^2\sigma^2}{N^3} + \frac{4\sigma}{N}$
$\mathsf{PCF2}_2 := (\Sigma^i =_1 X^j_\alpha) \wedge (\Theta^k =_1 X^l_\beta) \wedge (Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k) \wedge \overline{\mathsf{Bad}}$	$\frac{49q^2\sigma^2}{N^3} + \frac{\sigma}{N}$
$\mathsf{PCF2}_3 := (\Theta^i =_1 X^j_\alpha) \wedge (\Theta^k =_1 X^l_\beta) \wedge (Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k) \wedge \overline{\mathsf{Bad}}$	$\frac{49q^2\sigma^2}{N^3} + \frac{2\sigma}{N}$

To bound $\mathsf{PCF2}_1$, we first assume that $T^i = T^k$. As a matter of fact, the event $T^i = T^k$ induces $\Sigma^i =_1 \Sigma^k$ as $T^i = T^k$ implies $Y^j_\alpha = Y^l_\beta$ which implies $\Sigma^i =_1 \Sigma^k$. Therefore, one can write

$$\begin{aligned}
\Pr[\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}} \mid T^i = T^k] &= \Pr[\Sigma^i =_1 \Sigma^k \wedge \overline{\mathsf{Bad}} \mid T^i = T^k] \\
&\overset{[1]}{\leq} \sum_{i,k} \frac{4(\max\{l_i, l_k\} + 1)}{N} \leq \frac{4\sigma}{N},
\end{aligned} \tag{18}$$

where [1] follows from Claim 4.

Now, we do the analysis for the case $T^i \neq T^k$. Here, we assume that $M^l_\beta \neq M^k_\beta$ (the case for $M^l_\beta = M^k_\beta$ is similar and can be found in Appendix D.5). Let $\gamma \in \min \mathsf{NEQ}_{i,k}$. Note that γ cannot be equal to α and β simultaneously and wlog we assume that, $\gamma \neq \beta$. Moreover, since $\gamma \in \max\{l_i, l_j\}$, wlog we assume that, $\gamma \leq l_i$. Now, we write the three events (i.e. $\Sigma^i =_1 X^j_\alpha, \Sigma^k =_1 X^l_\beta$ and $Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k$) in terms of Y variables in the following matrix form:

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0/1} & X^j_\alpha \oplus b & \cdots \\ \mathbf{0} & \mathbf{0} & X^l_\beta \oplus b' & \cdots \\ \mathbf{0/1} & \mathbf{1} & T^i \oplus T^j & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_\gamma \\ Y^l_\beta \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Let us define the event $\mathsf{B} := (X^j_\alpha \oplus X^l_\beta \oplus b \oplus b' \oplus T^i \oplus T^k = \mathbf{0})$. It is easy to see that $\Pr[\mathsf{B}] \leq \frac{1}{N}$ as it induces a linear equation over Δ_0 and Δ_1. Now, it is easy to observe that, if $(A[1][2], A[3][1]) = (\mathbf{1}, \mathbf{1})$ and B holds, then $rank(A) \geq 2$, otherwise $rank(A) = 3$. Now, from the above observations and the Proposition 1 with our assumption $\sigma \leq \frac{N}{2}$, we have

$$\begin{aligned}
\Pr[\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}} \mid T^i \neq T^k] &\leq \Pr[\mathsf{PCF2}_1 \mid \overline{\mathsf{B}} \wedge T^i \neq T^k, \overline{\mathsf{Bad}}] \\
&\quad + \Pr[\mathsf{PCF2}_1 \mid \mathsf{B} \wedge \overline{\mathsf{Bad}} \wedge T^i \neq T^k] \cdot \Pr[\mathsf{B} \mid T^i \neq T^k] \\
&\leq \sum_{i,j,k,l} \sum_{\alpha,\beta} \frac{49}{N^3} \leq \frac{49q^2\sigma^2}{N^3}.
\end{aligned} \tag{19}$$

Now, combining both the cases we obtain

$$\begin{aligned}
\Pr[\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}}] &\leq \Pr[\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}} \mid T^i = T^k] + \Pr[\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}} \mid T^i \neq T^k] \\
&\leq \frac{4\sigma}{N} + \frac{49q^2\sigma^2}{N^3},
\end{aligned} \tag{20}$$

where [2] follows from Eqn. (18) and (19).

With a similar argument as above, one can show

$$\Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{49q^2\sigma^2}{N^3} + \frac{\sigma}{N}. \tag{21}$$

$$\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{Bad}}] \leq \frac{49q^2\sigma^2}{N^3} + \frac{2\sigma}{N}. \tag{22}$$

Proof of Eqn. (21) and (22) can be found in Appendix D.6 and D.7 respectively. Finally combining Eqn. (20), (21) and (22) we have,

$$\begin{aligned}\Pr[\mathsf{PCF2} \wedge \overline{\mathsf{Bad}}] &\leq \Pr[\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}}] + \Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{Bad}}] + \Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{Bad}}] \\ &\leq \frac{147q^2\sigma^2}{N^3} + \frac{7\sigma}{N}.\end{aligned}$$

8 Conclusion

We have presented a rate-1 single keyed block cipher based beyond birthday bound secure deterministic MAC. To the best of our knowledge, this is the first single keyed beyond birthday bound secure block cipher based variable input length PRF construction. Improving the PRF bound or giving a tight bound of the construction will be an interesting research problem. We believe, in a similar way, one can acheive the beyond birthday security of the single key variant of 3kf9, as proposed by Zhang et al. in ASIACRYPT, 2012. Moreover, our result on sum of permutation under conditional distribution could be applied in proving the security of single keyed construction that inherently uses the sum construction.

Acknowledgements

We would like to thank all the anonymous reviewers of FSE 2018 and CRYPTO 2017 for helping us in improving the work. Nilanjan Datta is supported by Information Security Education and Awareness (ISEA) project (Project Code. IIP). Liting Zhang is supported by National Key R & D Program of China (No. 2017YFB0802000) and National Natural Science Foundation of China (No. 61572484).

References

[Ber99] Daniel J. Bernstein. How to stretch random functions: The security of protected counter sums. *J. Cryptology*, 12(3):185–192, 1999.

[BGK99] Mihir Bellare, Oded Goldreich, and Hugo Krawczyk. Stateless evaluation of pseudorandom functions: Security beyond the birthday barrier. In *CRYPTO, 1999, Proceedings*, pages 270–287, 1999.

[BGR95] Mihir Bellare, Roch Guérin, and Phillip Rogaway. XOR macs: New methods for message authentication using finite pseudorandom functions. In *CRYPTO 1995, Proceedings*, pages 15–28, 1995.

[BI99] Mihir Bellare and Russell Impagliazzo. A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. *IACR Cryptology ePrint Archive*, 1999:24, 1999.

[BJK+16] Christof Beierle, Jérémy Jean, Stefan Kölbl, Gregor Leander, Amir Moradi, Thomas Peyrin, Yu Sasaki, Pascal Sasdrich, and Siang Meng Sim. The SKINNY family of block ciphers and its low-latency variant MANTIS. In *CRYPTO 2016, Proceedings, Part II*, pages 123–153, 2016.

[BKR00] Mihir Bellare, Joe Kilian, and Phillip Rogaway. The security of the cipher block chaining message authentication code. *J. Comput. Syst. Sci.*, 61(3):362–399, 2000.

[BR02] John Black and Phillip Rogaway. A block-cipher mode of operation for parallelizable message authentication. In *EUROCRYPT 2002*, pages 384–397, 2002.

[CLP14] Benoit Cogliati, Rodolphe Lampe, and Jacques Patarin. The indistinguishability of the XOR of k permutations. In *FSE 2014. Revised Selected Papers*, pages 285–302, 2014.

[CS16] Benoît Cogliati and Yannick Seurin. EWCDM: an efficient, beyond-birthday secure, nonce-misuse resistant MAC. In *CRYPTO 2016, Proceedings, Part I*, pages 121–149, 2016.

[DHT17] Wei Dai, Viet Tung Hoang, and Stefano Tessaro. Information-theoretic indistinguishability via the chi-squared method. In *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part III*, pages 497–523, 2017.

[DR00] Joan Daemen and Vincent Rijmen. Rijndael for AES. In *AES Candidate Conference*, pages 343–348, 2000.

[FLS+10] Niels Ferguson, Stefan Lucks, Bruce Schneier, Doug Whiting, Mihir Bellare, Tadayoshi Kohno, Jon Callas, and Jesse Walker. The skein hash function family. http://www.skein-hash.info/sites/default/files/skein1.3.pdf, 2010.

[GPR17] Peter Gazi, Krzysztof Pietrzak, and Michal Rybár. The exact security of PMAC. *IACR Cryptology ePrint Archive*, 2017:69, 2017.

[IK03] Tetsu Iwata and Kaoru Kurosawa. OMAC: one-key CBC MAC. In *Fast Software Encryption, 2003*, pages 129–153, 2003.

[JJV02] Éliane Jaulmes, Antoine Joux, and Frédéric Valette. On the security of randomized CBC-MAC beyond the birthday paradox limit: A new construction. In *Fast Software Encryption, 2002*, pages 237–251, 2002.

[JNP14] Jérémy Jean, Ivica Nikolic, and Thomas Peyrin. Tweaks and keys for block ciphers: The TWEAKEY framework. In *Advances in Cryptology - ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014, Proceedings, Part II*, pages 274–288, 2014.

[JTC11] JTC1. ISO/IEC 9797-1:2011 information technology - security techniques - message authentication codes (macs) - part 1: Mechanisms using a block cipher. 2011.

[LN17] Eik List and Mridul Nandi. Revisiting full-prf-secure PMAC and using it for beyond-birthday authenticated encryption. In *Topics in Cryptology - CT-RSA 2017 - The Cryptographers' Track at the RSA Conference 2017, San Francisco, CA, USA, February 14-17, 2017, Proceedings*, pages 258–274, 2017.

[LPTY16] Atul Luykx, Bart Preneel, Elmar Tischhauser, and Kan Yasuda. A MAC mode for lightweight block ciphers. *IACR Cryptology ePrint Archive*, 2016:190, 2016.

[Luc00] Stefan Lucks. The sum of prps is a secure PRF. In *EUROCRYPT 2000*, pages 470–484, 2000.

[Min10] Kazuhiko Minematsu. How to thwart birthday attacks against macs via small randomness. In *Fast Software Encryption, 2010*, pages 230–249, 2010.

[MM07] Kazuhiko Minematsu and Toshiyasu Matsushima. New bounds for pmac, tmac, and XCBC. In *FSE 2007, Revised Selected Papers*, pages 434–451, 2007.

[Nai15] Yusuke Naito. Full prf-secure message authentication code based on tweakable block cipher. In *Provable Security - 9th International Conference, ProvSec 2015, Kanazawa, Japan, November 24-26, 2015, Proceedings*, pages 167–182, 2015.

[Nan09] Mridul Nandi. Fast and secure cbc-type MAC algorithms. In *Fast Software Encryption, Revised Selected Papers*, pages 375–393, 2009.

[NIS05] NIST. Recommendation for block cipher modes of operation: The CMAC mode for authentication. *SP 800-38B*, 2005.

[NM08] Mridul Nandi and Avradip Mandal. Improved security analysis of PMAC. *J. Mathematical Cryptology*, 2(2):149–162, 2008.

[oS97] National Bureau of Standards. Data encryption standard. U.S. Departments of Commerce, FIPS, 1997.

[Pat08a] Jacques Patarin. The "coefficients h" technique. In *Selected Areas in Cryptography, 2008*, pages 328–345, 2008.

[Pat08b] Jacques Patarin. A proof of security in $o(2^n)$ for the xor of two random permutations. In *ICITS 2008*, pages 232–248, 2008.

[Pat10] Jacques Patarin. Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. *IACR Cryptology ePrint Archive*, 2010:287, 2010.

[Rog04] Phillip Rogaway. Efficient instantiations of tweakable blockciphers and refinements to modes OCB and PMAC. In *ASIACRYPT 2004, Proceedings*, pages 16–31, 2004.

[Sar10] Palash Sarkar. Pseudo-random functions and parallelizable modes of operations of a block cipher. *IEEE Trans. Information Theory*, 56(8):4025–4037, 2010.

[Yas10] Kan Yasuda. The sum of CBC macs is a secure PRF. In *CT-RSA 2010, Proceedings*, pages 366–381, 2010.

[Yas11] Kan Yasuda. A new variant of PMAC: beyond the birthday bound. In *CRYPTO 2011*, pages 596–609, 2011.

[ZWSW12] Liting Zhang, Wenling Wu, Han Sui, and Peng Wang. 3kf9: Enhancing 3gpp-mac beyond the birthday bound. In *ASIACRYPT 2012*, pages 296–312, 2012.

A Proof of Claim 3 from Section 6

Claim 3. Let M^i, M^j and M^k be three messages. Let $\alpha \in [l_j], \beta \in [l_k]$ and c be a non-zero constant. Then, for any $b, b' \in \{\mathbf{0}, \mathbf{1}\}$, we have

$$(a)\,\Pr[\Sigma^i = X^j_\alpha \oplus b \;|cX^j_\alpha = X^k_\beta \oplus b'] \;\; \leq \;\; \frac{2}{N}, \tag{23}$$

$$(b)\,\Pr[\Theta^i = X^j_\alpha \oplus b \;|cX^j_\alpha = X^k_\beta \oplus b'] \;\; \leq \;\; \frac{2}{N}, \tag{24}$$

where $\sigma \leq \frac{N}{2}$.

Proof. Let us consider a set $\Xi := \{(\delta_0, \delta_1) : (c2^\alpha \oplus 2^\beta)\delta_0 \oplus (c2^{2\alpha} \oplus 2^{2\beta})\delta_1 = c \cdot M^j_\alpha \oplus M^k_\beta \oplus b'\}$. We equivalently write Eqn. (23) and Eqn. (24) as follows

$$\Pr[\Sigma^i = X^j_\alpha \oplus b \;|(\Delta_0, \Delta_1) \in \Xi], \tag{25}$$

$$\Pr[\Theta^i = X^j_\alpha \oplus b \;|(\Delta_0, \Delta_1) \in \Xi]. \tag{26}$$

Now, for fixed indices $i \in [q], \alpha \in [l_i]$ and $b \in \{\mathbf{0}, \mathbf{1}\}$, we define the following set:

$$\mathcal{Z}^b_{i,\alpha} := \{(\delta_0, \delta_1) : 2^\alpha \delta_0 \oplus 2^{2\alpha}\delta_1 = M^i_\alpha \oplus b\}.$$

Moreover, we have $\mathcal{Z} := \cup_{b,i,\alpha} \mathcal{Z}^b_{i,\alpha}$. Now, It is easy to see that

$$(\Delta_0, \Delta_1) \in \Xi \Rightarrow (\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}. \tag{27}$$

In other words, if $(\Delta_0, \Delta_1) \in \mathcal{Z}$, then it cannot be the case that there exists some γ such that $c2^\alpha \oplus 2^\beta = 2^\gamma$ and $c2^{2\alpha} \oplus 2^{2\beta} = 2^{2\gamma}$ holds simultaneously.

Now, under the condition $(\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}$, we have the following:

$$(a)\,\Pr[\Sigma^i = X^j_\alpha \oplus b \mid (\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}] \;\; \leq \;\; \frac{2}{N}, \tag{28}$$

$$(b)\,\Pr[\Theta^i = X^j_\alpha \oplus b \mid (\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}] \;\; \leq \;\; \frac{2}{N}, \tag{29}$$

where we assume $\ell \leq \frac{N}{2}$. To justify Eqn. (28), we write the event $(\Sigma^i = X^j_\alpha \oplus b)$ as follows:

$$Y^i_1 \oplus \cdots \oplus Y^i_{l_i} = M^j_\alpha \oplus 2^\alpha.\Delta_0 \oplus 2^{2\alpha}.\Delta_1 \oplus b. \tag{30}$$

From the conditional event (i.e $(\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}$), we see that the right hand side of Eqn. (30) is non-zero which implies that the equation itself is non-trivial and hence rank of the equation is 1. Therefore, the result (a) follows from Proposition 1 assuming $\ell \leq \frac{N}{2}$.

Similarly, to justify Eqn. (29), we write the event $(\Theta^i = X^j_\alpha \oplus b)$ as follows:

$$2^{l_i}Y^i_1 \oplus \cdots \oplus 2Y^i_{l_i} = M^j_\alpha \oplus 2^\alpha.\Delta_0 \oplus 2^{2\alpha}.\Delta_1 \oplus b. \tag{31}$$

Like we argued before, since $(\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}$, it ensures the non-triviality of Eqn. (31) and hence, using Proposition 1, the result (b) follows assuming $\ell \leq \frac{N}{2}$.

Therefore, from Eqn. (25), Eqn. (27) and Eqn. (28) we have,

$$\Pr[\Sigma^i = X^j_\alpha \oplus b \;|cX^j_\alpha = X^k_\beta \oplus b'] \leq \Pr[\Sigma^i = X^j_\alpha \oplus b \mid (\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}] \leq \frac{2}{N}.$$

Similarly, from Eqn. (26), Eqn. (27) and Eqn. (29) we have,

$$\Pr[\Theta^i = X^j_\alpha \oplus b \;|cX^j_\alpha = X^k_\beta \oplus b'] \leq \Pr[\Theta^i = X^j_\alpha \oplus b \mid (\Delta_0, \Delta_1) \in \overline{\mathcal{Z}}] \leq \frac{2}{N}.$$

Hence, our result follows.

B Proof of Claim 4 from Section 7.1

Claim 4. *Let M^i and M^j be two distinct messages. If $\sigma \leq \frac{N}{2}$ then,*

$$(a)\,\Pr[\Sigma^i =_1 \Sigma^j \wedge \overline{\mathsf{Bad}}] \leq \frac{4(\max\{l_i, l_j\}+1)}{N} \quad \text{and} \quad (b)\,\Pr[\Theta^i =_1 \Theta^j \wedge \overline{\mathsf{Bad}}] \leq \frac{4}{N}.$$

Proof. To prove (a), let $\gamma \in \min \mathsf{NEQ}_{i,j}$ and $\gamma \neq \beta$. Moreover, $\gamma \in \max\{l_i, l_j\}$ and wlog we assume that, $\gamma \leq l_i$ and therefore Y^i_γ exists. Now, there are two possibilities:

Case (i): Let CollX_{ij} denotes the event that X^i_γ collides with any of $X^\star_\beta$ where $\star \in \{i, j\}$ and $\beta \in \max\{l_i, l_j\}$. For fixed i, j, we have

$$\Pr[\mathsf{CollX}_{ij}] \leq \frac{2\max\{l_i, l_j\}}{N},$$

which follows from Corollary 1. In this case, we bound the probability of the event (i.e. $\Sigma^i = \Sigma^j \oplus b \wedge \overline{\mathsf{Bad}}$) by probability of collision happens between X^i_γ and $X^\star_\beta$, where $b \in \{0,1\}$ be any fixed bit.

Case (ii): Let X^i_γ does not collide at all. In this case, X^i_γ is fresh and hence the equation induced by $\Sigma^i = \Sigma^j \oplus b$ is a non-trivial equation for the random variable Y^i_γ and hence the probability of the event will be bounded by $\frac{2}{N}$, which follows from Proposition 1 with the assumption that $\sigma \leq \frac{N}{2}$.

Therefore, we have

$$\begin{aligned} \Pr[\Sigma^i = \Sigma^j \oplus b \wedge \overline{\mathsf{l\text{-}Bad}}] &= \Pr[\Sigma^i = \Sigma^j \oplus b \mid \overline{\mathsf{l\text{-}Bad}} \wedge \overline{\mathsf{Coll}_{i,j}}] + \Pr[\mathsf{Coll}_{i,j}] \\ &\leq \frac{2}{N} + \frac{2\max\{l_i, l_j\}}{N} \\ &= \frac{2(\max\{l_i, l_j\}+1)}{N}. \end{aligned}$$

Now, we will prove (b). Here, we argue that given the condition $\overline{\mathsf{Bad}}$, the equation induced by $\Theta^i = \Theta^j \oplus b$ is always a non-trivial equation

$$(2^{l_i}Y^i_1 \oplus 2^{l_i-1}Y^i_2 \oplus \ldots \oplus 2Y^i_{l_i}) \oplus (2^{l_j}Y^j_1 \oplus 2^{l_j-1}Y^j_2 \oplus \ldots \oplus 2Y^j_{l_j}) = b. \qquad (32)$$

Case (i): $b = \mathbf{0}$. Let $l_i \neq l_j$ and wlog we assume that $l_i > l_j$. Clearly, the sum of all the coefficients of Eqn. (32) is $(2 \oplus \cdots 2^{l_i - l_j})$ which is non-zero as $l_i \leq \sigma < N$. When $l_i = l_k$, let $\alpha \in \mathsf{NEQ}_{ij}$, then the coefficient of Y^i_α is of the form $2^{l_i-\alpha+1}$ or $(2^{l_i-\alpha+1} \oplus 2^{l_i-\beta+1}$ (depending on whether X^i_α collides with X^*_β or not), which is non-zero. Hence, the equation is non-trivial.

Case (ii): $b = \mathbf{1}$. Here, the non-zero constant ensures Eqn. (32) to be non-trivial. Therefore, from Proposition 1, we obtain the result.

C Proof of Claim 5 from Section 7.2

Claim 5. *If $\overline{\mathsf{CollX}_{\mathbf{ijk}}}$ occurs, then the system of equations $\Sigma^i =_1 \Sigma^j$ and $\Theta^i =_1 \Theta^k$ has rank exactly 2.*

Proof. To prove this, we first rewrite the equations as $\Sigma_i = \Sigma_j \oplus b$ and $\Theta_i = \Theta_k \oplus b'$ where $b, b' \in \{0,1\}$. Now, we analyse the rank of these two simultaenous equations in case by case:

Case A $(\mathbf{b = 0, b' = 0})$. We analyze this case using the following subcases:

Case A.1 $(\mathbf{l_i = l_k})$. It is easy to see that in this case $|\mathsf{NEQ}_{i,k}| \geq 2$. We choose $\alpha \in \mathsf{NEQ}_{i,j} \cup \mathsf{ADD}_{i,j}$ and $\beta(\neq \alpha) \in \mathsf{NEQ}_{i,k}$. This is possible as $|\mathsf{NEQ}_{i,k}| \geq 2$. Now, consider the following two important observations: (i) The coefficient of the variable Y^i_α is $2^{l_i-\alpha}$ (if $Y^i_\alpha = Y^k_\alpha$) or $\mathbf{0}$ (else) for the event $(\Theta_i = \Theta_k)$, (ii) The coefficient of the variable Y^i_β is $\mathbf{0}$ (if $Y^i_\beta = Y^j_\beta$) or $\mathbf{1}$ (else) for the event $(\Sigma_i = \Sigma_j)$. So, the two equations can be written as

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0}\ /\ \mathbf{1} & \cdots \\ \mathbf{0}\ /\ 2^a & 2^b & \cdots \end{pmatrix}}_{A} \cdot \begin{pmatrix} Y^i_\alpha \\ Y^i_\beta \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $a = (l_i - \alpha + 1)$ and $b = (l_i - \beta + 1)$. As $\alpha \neq \beta$, $\det A \neq 0$ and *rank* of A is always 2.

Case A.2 ($\mathbf{l_i \neq l_k}$ and $\mathbf{M^j}$ is prefix of $\mathbf{M^i}$). As M^j is prefix of M^i, (i) $l_i \geq l_j$ and (ii) $Y^i_{l_j} = Y^j_{l_j}$. Now, it is easy to see that the coefficient of $Y^i_{l_i}$ and $Y^i_{l_j}$ in $(\Sigma^i = \Sigma^j)$ is $\mathbf{1}$ and $\mathbf{0}$ respectively. Moreover the coefficient of $Y^i_{l_j}$ in $(\Theta^i = \Theta^k)$ is $2^{l_i-l_j+1}$ (if $Y^i_{l_j} \neq Y^k_{l_j}$) or $2^{l_i-l_j+1} \oplus 2^{l_k-l_j+1}$ (else). Hence, the two equations can be written as

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0} & \cdots \\ \star & 2^a\ /\ (2^a \oplus 2^b) & \cdots \end{pmatrix}}_{A} \cdot \begin{pmatrix} Y^i_{l_i} \\ Y^i_{l_j} \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $a = (l_i - l_j + 1)$ and $b = (l_k - l_j + 1)$. As $l_i \neq l_k$, $\det A \neq 0$ and *rank* of A is always 2.

Case A.3 ($\mathbf{l_i \neq l_k}$ and $\mathbf{M^i}$ is prefix of $\mathbf{M^j}$). As M^j is prefix of M^i, (i) $l_i \geq l_j$ and (ii) $Y^i_{l_j} = Y^j_{l_j}$. Now, it is easy to see that the coefficient of $Y^i_{l_i}$ and $Y^i_{l_j}$ in $(\Sigma^i = \Sigma^j)$ is $\mathbf{1}$ and $\mathbf{0}$ respectively. Moreover the coefficient of $Y^i_{l_j}$ in $(\Theta^i = \Theta^k)$ is 2 (if $Y^i_{l_j} \neq Y^k_{l_j}$) or $2 \oplus 2^{l_k-l_j+1}$ (else). Hence, the two equations can be written as

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0} & \cdots \\ \star & 2\ /\ (2 \oplus 2^a) & \cdots \end{pmatrix}}_{A} \cdot \begin{pmatrix} Y^j_{l_j} \\ Y^i_{l_i} \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $a = (l_k - l_i + 1)$. As $l_i \neq l_k$, $\det A \neq 0$ and *rank* of A is always 2.

Case A.4 ($\mathbf{l_i \neq l_k}$ and $|\mathsf{NEQ}_{\mathbf{i,j}}| \geq \mathbf{1}$). Let $\alpha \in \mathsf{NEQ}_{i,j}$. Now, it is easy to see that the coefficient of both Y^i_α and Y^j_α in $(\Sigma^i = \Sigma^j)$ are $\mathbf{1}$. Moreover the coefficient of $Y^i_{l_j}$ in $(\Theta^i = \Theta^k)$ is $2^{l_i-l_j+1}$ (if $Y^i_{l_j} \neq Y^k_{l_j}$) or $2^{l_i-l_j+1} \oplus 2^{l_k-l_j+1}$ (else). Hence, the two equations can be written as

$$E. \begin{pmatrix} Y^i_\alpha \\ Y^j_\alpha \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where the matrix E is one of the following three: $\begin{pmatrix} \mathbf{1} & \mathbf{1} & \cdots \\ 2^{l_i-\alpha} & \mathbf{0} & \cdots \end{pmatrix}$, $\begin{pmatrix} \mathbf{1} & \mathbf{1} & \cdots \\ 2^{l_i-\alpha} & 2^{l_k-\alpha} & \cdots \end{pmatrix}$, $\begin{pmatrix} \mathbf{1} & \mathbf{1} & \cdots \\ 2^{l_i-\alpha} \oplus 2^{l_k-\alpha} & \mathbf{0} & \cdots \end{pmatrix}$. As $l_i \neq l_k$, the matrix E always has rank 2.

Case B. $(\mathbf{b = 0, b' = 1})$. Let $\alpha \in \mathsf{NEQ}_{i,j}$. So, the two equations can be written as

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0} & \cdots \\ \star & \mathbf{1} & \cdots \end{pmatrix}}_{A} \cdot \begin{pmatrix} Y^i_\alpha \\ \mathbf{1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

The rank of matrix A is always 2 as $\det A \neq 0$.

Case C. $(\mathbf{b}=\mathbf{1}, \mathbf{b}'=\mathbf{0})$**.** Let $\alpha \in \mathsf{NEQ}_{i,k}$. So, the two equations can be written as

$$\underbrace{\begin{pmatrix} \star & \mathbf{1} & \cdots \\ 2^{l_i-\alpha+1} & \mathbf{0} & \cdots \end{pmatrix}}_{A} \cdot \begin{pmatrix} Y^i_\alpha \\ \mathbf{1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

Again, the rank of matrix A is always 2 as $\det A \neq 0$.

Case D. $(\mathbf{b}=\mathbf{1}, \mathbf{b}'=\mathbf{1})$**.** Let $\alpha \in \mathsf{NEQ}_{i,k}$. So, the two equations can be written as

$$\underbrace{\begin{pmatrix} \mathbf{0}\ /\ \mathbf{1} & \mathbf{1} & \cdots \\ 2^{l_i-\alpha+1} & \mathbf{1} & \cdots \end{pmatrix}}_{A} \cdot \begin{pmatrix} Y^i_\alpha \\ \mathbf{1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

Clearly, rank of matrix A is always 2 as $\det A \neq 0$.

D Remaining Analysis for Bounding RCOLL, ECF, PCF1 and PCF2

In Sect. 7, we have skipped the detailed analysis of some bad cases due to similarities in their analysis with some other cases. For the sake of completeness, in this section, we provide the detailed proof of those cases.

D.1 Bounding Joint Probability of RCOLL_2 and $\overline{\mathsf{Bad}}$.

We bound the joint probability of the event RCOLL_1 and $\overline{\mathsf{Bad}}$ as follows:

$$\begin{aligned}
\Pr[\mathsf{RCOLL}_2, \overline{\mathsf{Bad}}] &= \sum_{i,j} \Pr[\Theta^i =_1 \Theta^j, \Sigma^i_{\mathsf{out}} \in \mathit{Ran}(\mathcal{L}_2), \overline{\mathsf{Bad}}] \\
&\overset{[1]}{=} \sum_{i,j} \Pr[\Theta^i =_1 \Theta^j, \overline{\mathsf{Bad}}] \cdot \Pr[\Sigma^i_{\mathsf{out}} \in \mathit{Ran}(\mathcal{L}_2)] \\
&\overset{[2]}{\leq} \sum_{i,j} \frac{4}{N} \cdot \frac{1}{N-(2q+\eta)} \\
&\leq \frac{4q^2}{N^2} \leq \frac{2\sigma}{N},
\end{aligned}$$

where [1] follows from the independence of the two events and [2] follows from Claim 4 and the maximum size of $\mathit{Ran}(\mathcal{L}_2)$ is $2q+\eta$. The last inequality follows from $q \leq \sigma \leq \frac{N}{2}$.

D.2 Bounding Joint Probability of ECF_2 and $\overline{\mathsf{Bad}}$.

Let $\gamma = \min \mathsf{NEQ}_{i,j}$. Clearly, $\gamma \leq \max\{l_i, l_j\}$ and wlog letus assume $l_i \geq l_k$. Now, we write the event ECF_2 (i.e. $\Sigma^i =_1 X^j_\alpha$ and $\Theta^i =_1 \Theta^k$) in terms of Y-variables and form the following set of equations in the following matrix form:

$$\underbrace{\begin{pmatrix} \mathbf{1} & X^j_\alpha \oplus b & \cdots \\ 2^{l_i-\gamma+1} & b' & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_\gamma \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Here we define B to be the event $(2^{l_i-\gamma+1}(X^j_\alpha \oplus b) = \mathbf{1})$ and perform the same analysis as done in for ECF_1 to obtain

$$\Pr[\mathsf{ECF}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{19q\sigma^2}{N^2}.$$

D.3 Bounding Joint Probability of ECF_3 and $\overline{\mathsf{Bad}}$.

Like the previous cases, we first write the event ECF_3 (i.e. $\Sigma^i =_1 X^j_\alpha$ and $\Theta^i =_1 X^k_\beta$) in terms of Y-variables and form the following set of equations in the following matrix form:

$$\underbrace{\begin{pmatrix} \mathbf{1} & X^j_\alpha \oplus b & \cdots \\ 2 & X^k_\beta \oplus b' & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_{l_i} \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Here we define B to be the event $(2X^j_\alpha \oplus X^k_\beta = 2b \oplus b')$ and perform the same analysis as before to obtain:

$$\Pr[\mathsf{ECF}_3 \wedge \overline{\mathsf{Bad}}] \leq \frac{19q\sigma^2}{N^2}.$$

D.4 Bounding Joint Probability of $\mathsf{PCF1}_2$ and $\overline{\mathsf{Bad}}$.

We represent the event $\mathsf{PCF1}_2$: $\Theta^i =_1 X^j_\alpha, Y^j_\alpha \oplus Y^k_\beta = T^i$ by Y-variables and form the following set of equations in the following matrix form:

$$\underbrace{\begin{pmatrix} 2 & X^j_\alpha \oplus b & \cdots \\ \mathbf{0}/\mathbf{1} & T^i & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_{l_i} \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

Now, we analyze this case exactly similar to PCF_1 by dividing it in two cases depending on whether $Y^i_{l_i} \neq Y^j_\alpha, Y^k_\beta$ or not. If $Y^i_{l_i} = Y^j_\alpha$ or $Y^i_{l_i} = Y^k_\beta$ then we similarly separate the case depending on whether $\mathsf{B} := X^j_\alpha = 2T^i \oplus b$ have occured or not and obtain the following

$$\Pr[\mathsf{PCF1}_2 \wedge \overline{\mathsf{Bad}}] \leq \frac{13q\sigma^2}{N^2}.$$

D.5 Bounding $\mathsf{PCF2}_1$ when $T^i \neq T^k$ and $M^l_\beta = M^k_\beta$

We represent the equations of $\mathsf{PCF2}_1$ when $T^i \neq T^k$ and $M^l_\beta = M^k_\beta$ in the following matrix form [8]:

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0}/\mathbf{1} & X^j_\alpha \oplus b & \cdots \\ \mathbf{0} & \mathbf{1} & X^l_\beta \oplus b' & \cdots \\ \mathbf{0}/\mathbf{1} & \mathbf{1} & T^i \oplus T^k & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_\gamma \\ Y^l_\beta \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Let $\mathsf{B}_1 := X^l_\beta = b' \oplus T^i \oplus T^k$, $\mathsf{B}_2 := X^j_\alpha \oplus X^l_\beta = b \oplus b' \oplus T^i \oplus T^k$ and $\mathsf{B}_3 := X^j_\alpha = b \oplus T^i \oplus T^k$. Now, consider the following observations:

[8] $M^l_\beta = M^k_\beta$ implies $A[2][2] = \mathbf{1}$.

- If $A[3][1] = \mathbf{0}$ then $\mathsf{B} := \mathsf{B}_1$ and $\det A[\cdot, 1..3]$ is $X^l_\beta \oplus b' \oplus T^i \oplus T^k$.
- If $(A[1][2], A[3][1]) = (\mathbf{0}, \mathbf{1})$, then $\mathsf{B} := \mathsf{B}_2$ and $\det A[\cdot, 1..3] = (X^j_\alpha \oplus X^l_\beta) \oplus (b \oplus b') \oplus (T^i \oplus T^k)$.
- If $(A[1][2], A[3][1]) = (\mathbf{1}, \mathbf{1})$, then $\mathsf{B} := \mathsf{B}_3$ and $\det A[\cdot, 1..3] = X^j_\alpha \oplus b \oplus T^i \oplus T^k$.
- As $\det \begin{pmatrix} \mathbf{1} & X^j_\alpha \oplus b \\ \mathbf{0} & X^l_\beta \oplus b' \end{pmatrix}$ is $X^l_\beta \oplus b' (\neq \mathbf{0})$, $rank(A) \geq 2$.

First of all, $\Pr[\mathsf{B}] \leq \frac{1}{N-1}$. Moreover, from the first three observations, it is clear that if $\overline{\mathsf{B}}$ occurs, then $\det A[\cdot, 1..3]$ is non-zero and hence $rank(A) = 3$. Hence, using Proposition 1 and the assumption $\sigma \leq \frac{N}{2}$, we have

$$\begin{aligned} \Pr[\mathsf{PCF2}_1 \wedge \overline{\mathsf{Bad}} \mid T^i \neq T^k] &\leq \Pr[\mathsf{PCF2}_1 \mid \overline{\mathsf{B}} \wedge T^i \neq T^k \wedge \overline{\mathsf{Bad}}] \\ &\quad + \Pr[\mathsf{PCF2}_1 \mid \mathsf{B} \wedge \overline{\mathsf{Bad}} \wedge T^i \neq T^k] \cdot \Pr[\mathsf{B} \mid T^i \neq T^k] \\ &\leq \sum_{i,j,k,l} \sum_{\alpha,\beta} \frac{49}{N^3} \leq \frac{49 q^2 \sigma^2}{N^3}. \end{aligned} \tag{33}$$

D.6 Bounding $\mathsf{PCF2}_2$: $\Sigma^i =_1 X^j_\alpha, \Theta^k =_1 X^l_\beta, Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k$

We follow the similar analysis as we did for bounding $\mathsf{PCF2}_1$. We first bound this event based on the occurence of $T^i \neq T^k$.

Case A: $T^i \neq T^k$**.** We analyse this case based on whether $M^l_\beta \neq M^k_\beta$ or $M^l_\beta = M^k_\beta$. We start with the assumption that $M^l_\beta \neq M^k_\beta$. Let $\gamma \in \min \mathsf{NEQ}_{i,k}$. Note that γ cannot be equal to α and β simultaneously and wlog we assume that $\gamma \neq \beta$. Moreover, since $\gamma \in \max\{l_i, l_j\}$, wlog we assume that $\gamma \leq l_i$. Now, we write the three events (i.e. $\Theta^i =_1 X^j_\alpha, \Theta^k =_1 X^l_\beta$ and $Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k$) in terms of Y variables and form the following set of equations in the following matrix form:

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0/1} & X^j_\alpha \oplus b & \cdots \\ \mathbf{0} & \mathbf{0} & X^l_\beta \oplus b' & \cdots \\ \mathbf{0/1} & \mathbf{1} & T^i \oplus T^k & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_\gamma \\ Y^l_\beta \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Note that this matrix is exactly equal to the matrix that we obtained in Case B of Sect. 7.4 while bounding $\mathsf{PCF2}_1$. Therefore, we directly have,

$$\begin{aligned} \Pr[\mathsf{PCF2}_2 \mid \overline{\mathsf{B}} \wedge \overline{\mathsf{l\text{-}Bad}} \wedge T^i \neq T^k] &= \Pr[A \cdot Y = \mathbf{0}] \overset{[1]}{\leq} \frac{32}{N^3}, \\ \Pr[\mathsf{PCF2}_2 \mid \mathsf{B} \wedge \overline{\mathsf{l\text{-}Bad}} \wedge T^i \neq T^k] &= \Pr[A \cdot Y = \mathbf{0}] \overset{[2]}{\leq} \frac{16}{N^2}, \end{aligned}$$

where [1] and [2] follows from Proposition 1, $\sigma \leq \frac{N}{2}$ and $\mathsf{B} := X^j_\alpha \oplus X^l_\beta = (b \oplus b') \oplus (T^i \oplus T^k)$. Therefore, combining these two cases we have,

$$\begin{aligned} \Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i \neq T^k] &\leq \Pr[\mathsf{PCF2}_2 \mid \overline{\mathsf{B}} \wedge T^i \neq T^k \wedge \overline{\mathsf{l\text{-}Bad}}] \\ &\quad + \Pr[\mathsf{PCF2}_2 \mid \mathsf{B} \wedge \overline{\mathsf{l\text{-}Bad}} \wedge T^i \neq T^k] \cdot \Pr[\mathsf{B} \mid T^i \neq T^k] \\ &\overset{[3]}{\leq} \frac{49}{N^3}. \end{aligned} \tag{34}$$

Now, we analyse the case when $M^l_\beta = M^k_\beta$. As before, we represent the equations of $\mathsf{PCF2}_2$ when $T^i \neq T^k$ and $M^l_\beta = M^k_\beta$ in the following matrix form:

$$\underbrace{\begin{pmatrix} \mathbf{1} & \mathbf{0/1} & X_\alpha^j \oplus b & \cdots \\ \mathbf{0} & 2^{l_k-\beta+1} & X_\beta^l \oplus b' & \cdots \\ \mathbf{0/1} & \mathbf{1} & T^i \oplus T^k & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y_\gamma^i \\ Y_\beta^l \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Let $\mathsf{B}_1 := X_\beta^l = 2^{l_k-\beta+1}(T^i \oplus T^k) \oplus b'$, $\mathsf{B}_2 := X_\beta^l \oplus 2^{l_k-\beta+1} \cdot (X_\alpha^j \oplus b) = 2^{l_k-\beta+1}(T^i \oplus T^k) \oplus b'$ and $\mathsf{B}_3 := 2^{l_k-\beta+1} \cdot (X_\alpha^j \oplus b) = 2^{l_k-\beta+1}(T^i \oplus T^k)$. Now,consider the following observations:

- If $A[3][1] = \mathbf{0}$ then we assign $\mathsf{B} := \mathsf{B}_1$ and we have $\det A[\cdot, 1..3]$ is $2^{l_k-\beta+1}(T^i \oplus T^k) \oplus (X_\beta^l \oplus b')$.
- If $(A[1][2], A[3][1]) = (\mathbf{0}, \mathbf{1})$, then we assign $\mathsf{B} := \mathsf{B}_2$ and we have $\det A[\cdot, 1..3] = 2^{l_k-\beta+1}(T^i \oplus T^k) \oplus (X_\beta^l \oplus b') \oplus 2^{l_k-\beta+1} \cdot (X_\alpha^j \oplus b)$.
- If $(A[1][2], A[3][1]) = (\mathbf{1}, \mathbf{1})$, then we assign $\mathsf{B} := \mathsf{B}_3$ and we have $\det A[\cdot, 1..3] = 2^{l_k-\beta+1}(T^i \oplus T^k)) \oplus 2^{l_k-\beta+1} \cdot (X_\alpha^j \oplus b)$.
- As $\det \begin{pmatrix} \mathbf{1} & X_\alpha^j \oplus b \\ \mathbf{0} & X_\beta^l \oplus b' \end{pmatrix}$ is $X_\beta^l \oplus b' (\neq \mathbf{0})$, $rank(A) \geq 2$.

From the first 3 observations, it is clear that in any case, if $\overline{\mathsf{B}}$ occurs, then $\det A[\cdot, 1..3]$ is non zero and hence $rank(A) = 3$. Hence we have,

$$\Pr[\mathsf{PCF2}_2 \mid \overline{\mathsf{B}} \wedge T^i \neq T^k \wedge \overline{\mathsf{I\text{-}Bad}}] = \Pr[A \cdot Y = \mathbf{0}] \overset{[1]}{\leq} \frac{32}{N^3},$$

where [1] follows from Proposition 1 and the assumption $\ell \leq \frac{N}{2}$. Moreover, from the last observation, we have

$$\Pr[\mathsf{PCF2}_2 \mid \mathsf{B} \wedge \overline{\mathsf{I\text{-}Bad}} \wedge T^i \neq T^k] = \Pr[A \cdot Y = \mathbf{0}] \overset{[2]}{\leq} \frac{16}{N^2},$$

where [2] follows from Proposition 1 and the assumption $\sigma \leq \frac{N}{2}$. Finally combining the above two cases we have

$$\begin{aligned} \Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i \neq T^k] &\leq \Pr[\mathsf{PCF2}_2 \mid \overline{\mathsf{B}} \wedge T^i \neq T^k \wedge \overline{\mathsf{I\text{-}Bad}}] \\ &\quad + \Pr[\mathsf{PCF2}_2 \mid \mathsf{B} \wedge \overline{\mathsf{I\text{-}Bad}} \wedge T^i \neq T^k] \cdot \Pr[\mathsf{B} \mid T^i \neq T^k] \\ &\overset{[3]}{\leq} \frac{49}{N^3}. \end{aligned} \tag{35}$$

Note that the conditional event of B conditioned on $T^i \oplus T^k \neq \mathbf{0}$ induces a linear equation over Δ_0 and Δ_1 and thus the probability of this individual event is bouded by $\frac{1}{N-1}$ which follows from Corollary 1.

Taking maximum probability bound of these two cases i.e. maximum bound of Eqn. 34 and Eqn. (35) we obtain,

$$\Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i \neq T^k] \leq \frac{49}{N^3}. \tag{36}$$

Case B: $T^i = T^k$. We analyse this case in the following two subcases:

Case B.1: $(\mathbf{b} = \mathbf{b}')$. Here, we observe that the event $T^i = T^k$ induces $\Sigma^i = \Theta^k$. Therefore, we have

$$\begin{aligned}\Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i = T^k] &= \Pr[\Sigma^i = \Theta^k \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i = T^k] \\ &\overset{[1]}{=} 0,\end{aligned}$$

[1] follows as we have separated the domain of collision points from the very beginning of our construction.

Case B.2: $(\mathbf{b} \neq \mathbf{b}')$. In this case $T^i = T^k$ induces $\Sigma^i = \Theta^k \oplus \mathbf{1}$. Therefore, we have

$$\begin{aligned}\Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i = T^k] &= \Pr[\Sigma^i = \Theta^k \oplus \mathbf{1} \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i = T^k] \\ &\overset{[1]}{=} \Pr[\Sigma^i = \Theta^k \oplus \mathbf{1} \wedge \overline{\mathsf{l\text{-}Bad}}] \overset{[2]}{\leq} \frac{2}{N},\end{aligned}$$

where [1] follows as the event is independent of the T^i values and [2] follows as the equation induced by the event $\Sigma^i = \Theta^k \oplus \mathbf{1}$ is a non-trivial equation and hence rank is 1 and the assumption $\sigma \leq \frac{N}{2}$.

Now combining the above two cases we have,

$$\Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i = T^k] \leq \frac{2}{N}. \tag{37}$$

Finally combining all the above cases we have,

$$\begin{aligned}\Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{l\text{-}Bad}}] &\leq \sum_{i,j,k,l} \sum_{\alpha,\beta} \Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i \neq T^k] \\ &\quad + \sum_{i,k} \Pr[\mathsf{PCF2}_2 \wedge \overline{\mathsf{l\text{-}Bad}} \mid T^i = T^k] \cdot \Pr[T^i = T^k] \\ &\leq \sum_{i,j,k,l} \sum_{\alpha,\beta} \frac{49}{N^3} + \sum_{i,k} \frac{2}{N^2} \quad \text{(From Eqn. (36) and Eqn. (37))} \\ &\leq \frac{49q^2\sigma^2}{N^3} + \frac{\sigma}{N} \ (\text{As } q \leq \sigma \leq \tfrac{N}{2}) \ . \end{aligned} \tag{38}$$

D.7 Bounding $\mathsf{PCF2}_3$: $\Theta^i =_1 X^j_\alpha, \Theta^k =_1 X^l_\beta, Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k$.

We follow the similar analysis as we did for bounding $\mathsf{PCF2}_1$ and PCF_2. As before, we first bound this event based on the occurence of $T^i \neq T^k$.

Case A: $T^i \neq T^k$. We analyse this case based on whether $M^l_\beta \neq M^k_\beta$ or $M^l_\beta = M^k_\beta$. We start with the assumption that $M^l_\beta \neq M^k_\beta$. Let $\gamma \in \min \mathsf{NEQ}_{i,k}$. Note that γ cannot be equal to α and β simultaneously and alog we assume that $\gamma \neq \beta$. Moreover, since $\gamma \in \max\{l_i, l_j\}$, wlog we assume that $\gamma \leq l_i$. Now, we write the three events (i.e. $\Theta^i =_1 X^j_\alpha, \Theta^k =_1 X^l_\beta$ and $Y^j_\alpha \oplus Y^l_\beta = T^i \oplus T^k$) in terms of Y variables and form the following set of equations in the following matrix form: [9]

$$\underbrace{\begin{pmatrix} 2^{l_i-\gamma+1} & \mathbf{0}/2^{l_i-\beta+1} & X^j_\alpha \oplus b & \cdots \\ \mathbf{0} & \mathbf{0} & X^l_\beta \oplus b' & \cdots \\ \mathbf{0/1} & \mathbf{1} & T^i \oplus T^k & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_\gamma \\ Y^l_\beta \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

[9] In the matrix A, $A[1][2]$ can't be anything other than $\mathbf{0}$ or $2^{l_i-\beta+1}$ due to $\overline{\mathsf{Bad}_2}$ condition.

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Now, we consider the following three observations:

- If $(A[1][2], A[3][1]) \neq (2^{l_i-\beta+1}, \mathbf{1})$ then $\det A[\cdot, 1..3]$ is $2^{l_i-\gamma+1}.(X^l_\beta \oplus b')(\neq \mathbf{0})$, implying $rank(A) = 3$.
- If $(A[1][2], A[3][1]) = (2^{l_i-\beta+1}, \mathbf{1})$ then $\det A[\cdot, 1..3]$ is $(2^{l_i-\gamma+1} \oplus 2^{l_i-\beta+1}) \cdot (X^l_\beta \oplus b')$, which is non zero as $\gamma \neq \beta$ and $X^l_\beta \neq b'$. Hence, $rank(A) = 3$.

Clearly from the above observations, we have

$$\Pr[\mathsf{PCF2}_3 \mid T^i \neq T^k \wedge \overline{\mathsf{I\text{-}Bad}}] \quad = \quad \Pr[A \cdot Y = \mathbf{0}] \overset{[1]}{\leq} \frac{32}{N^3},$$

where [1] follows from Proposition 1 and the assumption $\ell \leq \frac{N}{2}$. Therefore, we have

$$\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i \neq T^k] \leq \Pr[\mathsf{PCF2}_3 \mid T^i \neq T^k \wedge \overline{\mathsf{I\text{-}Bad}}] \leq \frac{32}{N^3}. \tag{39}$$

Now, we analyse the case when $M^l_\beta = M^k_\beta$. As before, we represent the equations of $\mathsf{PCF2}_3$ when $T^i \neq T^k$ and $M^l_\beta = M^k_\beta$ in the following matrix form:

$$\underbrace{\begin{pmatrix} 2^{l_i-\gamma+1} & \mathbf{0}/2^{l_i-\beta+1} & X^j_\alpha \oplus b & \cdots \\ \mathbf{0} & 2^{l_k-\beta+1} & X^l_\beta \oplus b' & \cdots \\ \mathbf{0}/\mathbf{1} & \mathbf{1} & T^i \oplus T^k & \cdots \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} Y^i_\gamma \\ Y^l_\beta \\ \mathbf{1} \\ \vdots \end{pmatrix}}_{Y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $b, b' \in \{\mathbf{0}, \mathbf{1}\}$. Let $\mathsf{B}_1 := X^l_\beta = 2^{l_k-\beta+1}(T^i \oplus T^k) \oplus b'$. $\mathsf{B}_2 := 2^{l_i-\gamma+1} \cdot (X^l_\beta \oplus b') \oplus 2^{l_k-\beta+1} \cdot (X^j_\alpha \oplus b) = 2^{l_i+l_k-(\gamma+\beta)+2}(T^i \oplus T^k)$ and $\mathsf{B}_3 := 2^{l_i+1} \cdot (X^l_\beta \oplus b') \cdot (2^{-\gamma} \oplus 2^{-\beta}) \oplus 2^{l_k-\beta+1} \cdot (X^j_\alpha \oplus b) = 2^{l_i+l_k-(\gamma+\beta)+2}(T^i \oplus T^k)$. Again, consider the following observations:

- If $A[3][1] = \mathbf{0}$ then we assign $\mathsf{B} := \mathsf{B}_1$ and we have $\det A[\cdot, 1..3]$ is $2^{l_i-\gamma+1}.(X^l_\beta \oplus b' \oplus 2^{l_k-\beta+1}(T^i \oplus T^k))$.
- If $(A[1][2], A[3][1]) = (\mathbf{0}, \mathbf{1})$, then we assign $\mathsf{B} := \mathsf{B}_2$ and we have $\det A[\cdot, 1..3] = 2^{l_i-\gamma+1}.(X^l_\beta \oplus b' \oplus 2^{l_k-\beta+1}(T^i \oplus T^k)) \oplus 2^{l_k-\beta+1} \cdot (X^j_\alpha \oplus b)$.
- If $(A[1][2], A[3][1]) = (2^{l_i-\beta+1}, \mathbf{1})$, then we assign $\mathsf{B} := \mathsf{B}_3$ and we have $\det A[\cdot, 1..3] = 2^{l_i-\gamma+1} \cdot (X^l_\beta \oplus b' \oplus 2^{l_k-\beta+1}(T^i \oplus T^k)) \oplus 2^{l_k-\beta+1} \cdot (X^j_\alpha \oplus b) \oplus 2^{l_i-\beta+1} \cdot (X^l_\beta \oplus b')$.
- As $\det \begin{pmatrix} 2^{l_i-\gamma+1} & X^j_\alpha \oplus b \\ \mathbf{0} & X^l_\beta \oplus b' \end{pmatrix}$ is $2^{l_i-\gamma+1} \cdot (X^l_\beta \oplus b')(\neq \mathbf{0})$, $rank(A) \geq 2$.

From the first 3 observations, it is clear that in each case, if $\overline{\mathsf{B}}$ occurs, then $\det A[\cdot, 1..3]$ is non zero and hence $rank(A) = 3$. Hence we have,

$$\Pr[\mathsf{PCF2}_3 \mid \overline{\mathsf{B}} \wedge T^i \neq T^k \wedge \overline{\mathsf{I\text{-}Bad}}] = \Pr[A \cdot Y = \mathbf{0}] \overset{[1]}{\leq} \frac{32}{N^3},$$

where [1] follows from Proposition 1 and the assumption $\sigma \leq \frac{N}{2}$. Moreover, from the last observation, we have

$$\Pr[\mathsf{PCF2}_3 \mid \mathsf{B} \wedge \overline{\mathsf{I\text{-}Bad}} \wedge T^i \neq T^k] = \Pr[A \cdot \widetilde{Y} = \mathbf{0}] \overset{[2]}{\leq} \frac{16}{N^2},$$

where [2] follows from Proposition 1 and the assumption $\ell \leq \frac{N}{2}$. Finally combining the above two cases we have

$$\begin{aligned}\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i \neq T^k] &\leq \Pr[\mathsf{PCF2}_3 \mid \overline{\mathsf{B}} \wedge T^i \neq T^k \wedge \overline{\mathsf{I\text{-}Bad}}] \\ &\quad + \Pr[\mathsf{PCF2}_3 \mid \mathsf{B} \wedge \overline{\mathsf{I\text{-}Bad}} \wedge T^i \neq T^k] \cdot \Pr[\mathsf{B} \mid T^i \neq T^k] \\ &\overset{[3]}{\leq} \frac{49}{N^3}. \qquad (40)\end{aligned}$$

Note that the conditional event of B conditioned on $T^i \oplus T^k \neq \mathbf{0}$ induces a linear equation over Δ_0 and Δ_1 and thus the probability of this individual event is bouded by $\frac{1}{N}$, which follows from Corollary 1.

Taking maximum probability bound of these two cases i.e. maximum bound of Eqn. 39 and Eqn. (40) we obtain,

$$\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i \neq T^k] \leq \frac{49}{N^3}. \qquad (41)$$

Case B: $T^i = T^k$. We analyze this case in the following two subcases:

Case B.1: $(\mathbf{b} = \mathbf{b}')$. Here we observe that the event $T^i = T^k$ induces $\Theta^i = \Theta^k$. Therefore, we have

$$\begin{aligned}\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i = T^k] &= \Pr[\Theta^i = \Theta^k \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i = T^k] \\ &\overset{[1]}{=} \Pr[\Theta^i = \Theta^k \wedge \overline{\mathsf{I\text{-}Bad}}] \overset{[2]}{\leq} \frac{4}{N},\end{aligned}$$

where [1] follows similar to previous analysis and [2] follows from the proof of Claim 4 and $\sigma \leq \frac{N}{2}$.

Case B.2: $(\mathbf{b} \neq \mathbf{b}')$. Here we observe that the event $T^i = T^k$ induces $\Theta^i = \Theta^k \oplus \mathbf{1}$. Therefore, we have

$$\begin{aligned}\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i = T^k] &= \Pr[\Theta^i = \Theta^k \oplus \mathbf{1} \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i = T^k] \\ &\overset{[1]}{=} \Pr[\Theta^i = \Theta^k \oplus \mathbf{1} \wedge \overline{\mathsf{I\text{-}Bad}}] \overset{[2]}{\leq} \frac{4}{N},\end{aligned}$$

where [1] follows the same argument as in Case B.1 and [2] follows from Claim 4 and the assumption $\sigma \leq \frac{N}{2}$.

Hence, we have

$$\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i = T^k] \leq \frac{4}{N}. \qquad (42)$$

Finally combining all the above cases we have,

$$\begin{aligned}\Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}}] &\leq \sum_{i,j,k,l} \sum_{\alpha,\beta} \Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i \neq T^k] \\ &\quad + \sum_{i,k} \Pr[\mathsf{PCF2}_3 \wedge \overline{\mathsf{I\text{-}Bad}} \mid T^i = T^k] \cdot \Pr[T^i = T^k] \\ &\leq \sum_{i,j,k,l} \sum_{\alpha,\beta} \frac{49}{N^3} + \sum_{i,k} \frac{4}{N} \cdot \frac{1}{N} \quad \text{(From Eqn. (41) and Eqn. (42))} \\ &\leq \frac{49q^2\sigma^2}{N^3} + \frac{2\sigma}{N}. \qquad (43)\end{aligned}$$