On The Exact Security of Message Authentication Using Pseudorandom Functions

Ashwin Jha[1], Avradip Mandal[2] and Mridul Nandi[1]

[1] Indian Statistical Institute, Kolkata, India, {ashwin.jha1991,mridul.nandi}@gmail.com
[2] Fujitsu Laboratories Of America, Sunnyvale, USA, avradip@gmail.com

Abstract. Traditionally, modes of Message Authentication Codes(MAC) such as Cipher Block Chaining (CBC) are instantiated using block ciphers or keyed Pseudo Random Permutations(PRP). However, one can also use domain preserving keyed Pseudo Random Functions(PRF) to instantiate MAC modes. The very first security proof of CBC-MAC [BKR00], essentially modeled the PRP as a PRF. Until now very little work has been done to investigate the difference between PRP vs PRF instantiations. Only known result is the rather loose folklore PRP-PRF transition of any PRP based security proof, which looses a factor of $O(\frac{\sigma^2}{2^n})$ (domain of PRF/PRP is $\{0,1\}^n$ and adversary makes σ many PRP/PRF calls in total). This loss is significant, considering the fact tight $\Theta(\frac{q^2}{2^n})$ security bounds have been known for PRP based EMAC and ECBC constructions (where q is the total number of adversary queries). In this work, we show for many variations of encrypted CBC MACs (i.e. EMAC, ECBC, FCBC, XCBC and TCBC), random function based instantiation has a security bound $O(\frac{q\sigma}{2^n})$. This is a significant improvement over the folklore PRP/PRF transition. We also show this bound is optimal by providing an attack against the underlying PRF based CBC construction. This shows for EMAC, ECBC and FCBC, PRP instantiations are substantially more secure than PRF instantiations. Where as, for XCBC and TMAC, PRP instantiations are at least as secure as PRF instantiations.

Keywords: MAC · CBC · EMAC · XCBC · FCBC · TMAC · domain preserving PRF · PRP

1 Introduction

Message Authentication Codes or MACs are indispensable symmetric key cryptographic primitives for providing communication integrity. Cipher Block Chaining MAC or CBC-MAC is a popular mode of operation for constructing a MAC from a block cipher. The CBC mode of operation was part of ISO standard [ISO11].

In prior works, the security of CBC-MAC has been extensively studied [BKR00, BPR05]. However, one drawback of CBC-MAC is all the messages must come from a prefix-free family. To circumvent this issue EMAC or Encrypted MAC was developed as part of RACE project [BdBB+95] and later it was proven secure by Petrank and Rackoff [PR00]. Afterwards, in [Pie06, JN16] the security bounds was improved and it was shown to be optimal. However, still EMAC is only able to handle messages that are exact multiple of the block length (input size of the underlying block cipher). Certainly, one can use 10^* padding to convert the input message length to multiple of block size. This padded EMAC is known as EMAC*, which suffers from the drawback - even when the message is exact multiple of block length one extra block cipher call is required. In [BR00, BR05], Black and Rogaway proposed three refinements to EMAC called ECBC, FCBC and XCBC that can handle arbitrary length messages without the extra block cipher call for messages which are exact multiples of block length. Black and Rogaway's proposals (as well as

EMAC) use multiple keys. In [KI03, KI04] and [IK03a] Iwata and Kurosawa modified the XCBC construction and called it TMAC and OMAC, such that it can be instantiated with only two or one key respectively, where as retaining all its previous benefits. Later, its security bound was improved in [IK03b, Nan09, MM07].

In all these previous works the underlying primitive was assumed to be a block cipher or a keyed family of random permutations. However, the CBC mode of operation can also be used to construct a MAC when the underlying primitive is a domain preserving pseudo random function (PRF) or a keyed family of domain preserving random functions. In fact, the very first security proof of CBC-MAC in [BKR00] essentially models the block cipher as a PRF. We believe, when it comes to CBC-MAC the choice of block ciphers or pseudo random permutations (PRPs) over PRFs as the underlying primitive is primarily for historical reasons, i.e. fast, secure, standardized block cipher implementations are readily available compared to pseudo random functions. In fact, there has been little study about the security of CBC-MACs instantiated by PRFs. Note, any PRP based MAC security bound can be transformed to PRF based MAC security bound with a loss of $O(\sigma^2/2^n)$ security, where σ is the total number PRP/PRF calls in all queries with n-bits block length. However this generic transformation is rather loose, and may not necessarily capture the exact security of PRF based instantiations.

We only came across three prior works. First work is a study of iterated MAC constructions by Preneel and van Oorschot [PvO99], which claims tight bounds for PRF-based iterated MACs. But, we observe that the proof of the concerned result [PvO99, Lemma 2] has a flaw (see Section 3.2.1). In particular, their argument only works when the inputs to the underlying PRF are distinct at each message block index. This is certainly not true for CBC-MACs as it is possible that two internal inputs collide for a given message. Second one is the Diploma Thesis of Robert Berke [Ber03] where he showed, for prefix free messages CBC-MAC can be attacked with advantage $\Omega(\frac{\ell^2 q^2}{2^n})$ (here the adversary makes q many MAC queries of length at most ℓ blocks and block length is n-bits). This is in fact a non trivial result which shows the gap between PRP vs PRF instantiation, because for PRP instantiation of CBC-MACs such attacks are ruled out by Bellare et al. [BPR05] which shows the upper bound is actually $O(\frac{\ell q^2}{2^n})$. The third work is Guo et al's [GJMN15] recent Collision Distinguisher for an Iterated Random Function with advantage $\Omega(\frac{\ell q^2}{2^n})$. As it turns out this attack can be easily translated to attack against PRF based EMAC, ECBC, FCBC, XCBC and TMAC.

Given a lack of understanding of exact security of PRF based CBC-MAC constructions, the primary goal of this work is to understand the gap between PRP vs PRF instantiation of CBC-MAC and its derivative constructions. In Table 1, we provide best known (prior to this work) security bounds for PRP and PRF based CBC-MAC, EMAC constructions. Note, the best known attack (lower bound) for all PRP based constructions is the trivial birthday attack. We provide our improved bounds covered in this work in Table 2.

1.1 Our Contribution

Theorem 4 summarizes the main technical contribution of this work. We show PRF based EMAC constructions (EMAC, ECBC, FCBC, XCBC, TMAC) has an optimal security bound $\Theta(\frac{q\sigma}{2^n})$. The previously known best upper bound was $O(\frac{\sigma^2}{2^n})$ (a straight forward extension of PRP based bounds from [BPR05, Pie06, JN16, IK03b], by bounding PRP-PRF advantage). Note, for the PRP based EMAC and ECBC the upper bound is known to be optimal $O(\frac{q^2}{2^n})$. In fact we show our bound is optimal by demonstrating an attack against underlying PRF based CBC-MAC construction with advantage $\Omega(\frac{\ell q^2}{2^n})$, which becomes $\Omega(\frac{q\sigma}{2^n})$ for $\ell = \sigma/q$ (Theorem 3).

As we discussed before, Guo et al's recent work [GJMN15] also shows a $\Omega(\frac{\ell q^2}{2^n})$ attack

Table 1: Previously known security bounds of CBC-MACs

	Random Permutation		Random Function	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
CBC-MAC (Equal Length)	$\Omega(\frac{q^2}{2^n})$	$O(\frac{\ell q^2}{2^n})$ [BPR05]	$\Omega(\frac{q^2}{2^n})$	$O(\frac{\ell q^2+\sigma^2}{2^n})$ [a]
CBC-MAC (Prefix Free)	$\Omega(\frac{q^2}{2^n})$	$O(\frac{\ell q^2}{2^n})$ [BPR05]	$\Omega(\frac{\ell^2 q^2}{2^n})$ [Ber03]	$O(\frac{\ell q^2+\sigma^2}{2^n})$ [a]
EMAC, ECBC,FCBC	$\Omega(\frac{q^2}{2^n})$	$O(\frac{q^2}{2^n})$ [Pie06, JN16]	$\Omega(\frac{\ell q^2}{2^n})$ [GJMN15][b]	$O(\frac{\sigma^2}{2^n})$ [a]
XCBC, TMAC	$\Omega(\frac{q^2}{2^n})$	$O(\frac{\sigma^2}{2^n})$ [IK03b], $O(\frac{\ell q^2}{2^n})$ [MM07][c]	$\Omega(\frac{\ell q^2}{2^n})$ [GJMN15][b]	$O(\frac{\sigma^2}{2^n})$ [a]

[a] Trivial extension of Permutation based bound. All together there are σ many random function calls.
[b] Subsection 1.1 talks about the limitation of this bound.
[c] If the attacker makes q queries of block length $\ell_1, \cdots, \ell_q$, then $\sigma = \sum_{i=1}^{q} \ell_i$. If one query is substantially longer than other queries then $\sigma \ll \ell q$ and $\sigma^2 < \ell q^2$. So σ^2 and ℓq^2 are incomparable.

Table 2: New security bounds for CBC-MACs proved in this paper.

	Random Function	
	Lower Bound	Upper Bound
CBC-MAC (Equal Length)	$\Omega(\frac{q\sigma}{2^n})$	-
EMAC, ECBC	$\Omega(\frac{q\sigma}{2^n})$ [a]	$O(\frac{q\sigma}{2^n})$
FCBC	$\Omega(\frac{q\sigma}{2^n})$ [a]	$O(\frac{q\sigma}{2^n})$
XCBC, TMAC	$\Omega(\frac{q\sigma}{2^n})$ [a]	$O(\frac{q\sigma}{2^n})$

[a] See Subsection 1.1, why our bound is a substantial improvement over [GJMN15]

against PRF based CBC-MAC construction. However, [GJMN15] can only achieve the bound when $q^2 \leq \frac{2^n}{\ell^2}$. Subject to this condition, $\Omega(\frac{\ell q^2}{2^n})$ can be some constant only when q is $\Omega(2^{n/2})$. This lower bound on q is actually no better than the trivial birthday attack with advantage $\Omega(\frac{q^2}{2^n})$. Compared to that in Theorem 3 we show our lower bound holds as long as $\ell < \min(\frac{2^n}{5184}, \frac{2^{\frac{n}{2}}}{4\sqrt{3}}, \frac{2^{\frac{n}{3}}}{\sqrt[3]{36}})$. In fact, we can choose $\ell, q \in \Theta(2^{n/3})$ to achieve a constant advantage.

Moreover, our attack against PRF based EMAC in Section 6, actually induces a collision event against PRF based CBC-MAC construction. In our attack all the messages are of equal length. Previously, the best known lower bound against PRF based CBC-MAC for equal length messages was actually the trivial birthday attack. Whether this is a tight bound for Equal length PRF based CBC-MACs or not, is still an open problem. Note, for Prefix Free PRF based-MAC construction, Berke [Ber03] showed $\Omega(\frac{\ell^2 q^2}{2^n})$ attack matching the known upper bound. But that attack can not be extended to Equal Length message case.

2 Basic Definitions and Notations

2.1 Basic Notation

We use lowercase letters such as i, j for indices and integers, and f, g, h as functions. In particular, we fix a positive integer n. For notational simplicity we denote 2^n by N. For any two integers $a \leq b$, we write $[a..b]$ (or simply $[b]$ when $a = 1$) to denote the set $\{a, a+1, \ldots, b\}$. We use uppercase letters X, Y for variables, and calligraphic uppercase

letters $\mathcal{S}, \mathcal{T}$ for sets. Let $\sum := \{0,1\}$, and $\mathcal{B} := \sum^n$, where the elements of $\sum$ and $\mathcal{B}$ are called **bits** and **blocks** respectively. Let ϕ be a property defined for the elements of $\mathcal{S}$ then $\mathcal{S}[\phi]$ denotes the subset

$$\{X \in \mathcal{S} \mid X \text{ satisfies } \phi\}.$$

For any set $\mathcal{S}$, let $\mathcal{S}^{\leq \ell} := \cup_{i=1}^{\ell} \mathcal{S}^i$, $\mathcal{S}^+ := \cup_{i=1}^{\infty} \mathcal{S}^i$, and $\mathcal{S}^* := \cup_{i=0}^{\infty} \mathcal{S}^i$, where $\mathcal{S}^0 = \emptyset$, the empty set.

For a finite set $\mathcal{S}$, $X \xleftarrow{\$} \mathcal{S}$ denotes the uniform random sampling of X from $\mathcal{S}$. For any function $f : \mathcal{B} \to \mathcal{B}$, $f^{(0)}(X) = X$ and $f^{(i)}(X) = f \circ f^{(i-1)}(X)$ for $i \geq 1$. Let Func be the set of all functions from $\mathcal{B}$ to $\mathcal{B}$. We define set $\mathsf{Func}_\perp$ as

$$\mathsf{Func}_\perp := \{f \mid f : \mathcal{B} \cup \{\perp\} \to \mathcal{B} \wedge f(\perp) = 0^n\}.$$

Note that the uniform distribution over $\mathsf{Func}_\perp$ has the same probability mass function (p.m.f) as Func, i.e., a constant function taking up the value N^{-N}.

2.2 Collision Probability

We denote the probability of having a collision on k elements chosen uniformly and independently from $\mathcal{B}$ as $\mathbf{cp}_k$. It is well known that,

$$1 - \exp^{\frac{-k(k-1)}{2N}} \leq \mathbf{cp}_k \leq \frac{k(k-1)}{2N} \text{ [BG08]}. \tag{1}$$

2.3 Notation on Sequences and Strings

Let $\mathcal{I}$ and $\mathcal{S}$ be two sets. An $\mathcal{S}$ sequence X over the index set $\mathcal{I}$ is denoted as $(X_\alpha)_{\alpha \in \mathcal{I}}$ where $X_\alpha \in \mathcal{S}$ for all $\alpha \in \mathcal{I}$. In this paper we mostly consider *block sequences*, i.e. $\mathcal{S} = \mathcal{B}$ and *bit sequences*, i.e. $\mathcal{S} = \sum$. Length of the sequence, denoted by $|X|_{|\mathcal{S}|}$ (or simply $|X|$ when $\mathcal{S} = \sum$) is $|\mathcal{I}|$, the size of the index set. When the index set is $[a..b]$, we also write the sequence as a tuple or vector $X_{[a..b]} := (X_a, \ldots, X_b)$ ($X_{[b]}$ when $a = 1$). Sometimes, by abusing notation, X also represents the set $\{X_\alpha : \alpha \in \mathcal{I}\}$. Similarly $X_{[a..b]}$ represents $\{X_\alpha : \alpha \in [a..b]\}$. Note that, we can view $\mathcal{S}^{\leq \ell}$ as the set of all $\mathcal{S}$ sequences of lengths at most ℓ; $\mathcal{S}^+$ as the set of all $\mathcal{S}$ sequences of positive lengths; and $\mathcal{S}^*$ as the set of all $\mathcal{S}$ sequences (including zero length sequence). We may also view a sequence as a string. For a string $X = A \| B$, A (respective. B) is said to be a *prefix* (respective. *suffix*) of X. We write $A <_p X$ if A is a prefix of X and $B <_s X$ if B is a suffix of X. For two strings A and B of lengths a and b respectively, a non-negative integer $p := \mathsf{LCP}(A; B)$ (respective. $s := \mathsf{LCS}(A; B)$) is called the *largest common prefix* (respective. *largest common suffix*), if $A_{[1..p]} = B_{[1..p]}$ and $A_{[p+1]} \neq B_{[p+1]}$ (respective. $A_{[a-s+1..a]} = B_{[b-s+1..b]}$ and $A_{[a-s]} \neq B_{[b-s]}$).

Definition 1. Given a sequence X with an index set $\mathcal{I}$, we associate an equivalence relation $\sim_X$ over $\mathcal{I}$ as follows: $\alpha \sim_X \beta$ if $X_\alpha = X_\beta$.

Let $f : \mathcal{D} \to \mathcal{R}$ and let X and Y be $\mathcal{D}$ and $\mathcal{R}$ sequences respectively with same index set $\mathcal{I}$. We write $X \xmapsto{f} Y$ to mean that $f(X_\alpha) = Y_\alpha$ for all $\alpha \in \mathcal{I}$ and we simply say that f *multi-maps* X *to* Y. This is a property of a function. So $\mathsf{Func}[X \longmapsto Y]$represents the set of all functions f multi-mapping X to Y. We say that (X, Y) is *function compatible* if there exists a function f such that $X \xmapsto{f} Y$. The following proposition is easily verifiable.

Proposition 1. *Let x and y be $\mathcal{D}$ and $\mathcal{R}$ sequences over an index set $\mathcal{I}$. Then (x, y) is function compatible if $\sim_x \subseteq \sim_y$.*[1]

[1]Here we view the equivalence relation, $\sim_x$ as $\{(\alpha, \beta) \in \mathcal{I} \times \mathcal{I} : \alpha \sim_x \beta\} \subseteq \mathcal{I} \times \mathcal{I}$. The equivalence relation $\sim_y$ is viewed similarly.

2.4 PRF Security of Keyed Functions

Let $\mathcal{D} \subseteq \mathcal{B}^+$ be a finite set. Let $\mathsf{Func}(\mathcal{D}, \mathcal{B})$ be the set of all functions from $\mathcal{D}$ to $\mathcal{B}$. A **Random Function** from $\mathcal{D}$ to $\mathcal{B}$ is $\mathcal{F}(\mathcal{D}, \mathcal{B}) \xleftarrow{\$} \mathsf{Func}(\mathcal{D}, \mathcal{B})$. A **Keyed Function Family** $F_{\mathcal{K}}(\mathcal{D}, \mathcal{B})$ with key $K \in \mathcal{K}$ is a function from $\mathcal{K} \times \mathcal{D}$ to $\mathcal{B}$. $F_K := F(K, \cdot)$ is called a **keyed function** from $\mathcal{D}$ to $\mathcal{B}$. When the domain and range sets are understood, we denote random function as $\mathcal{F}$ and keyed function as F_K.

Definition 2. Let $F_{\mathcal{K}}$ be a keyed function family from $\mathcal{D}$ to $\mathcal{B}$. We define the **PRF-advantage** (or pseudorandom function advantage) of an adversary $\mathbf{A}$ against $F_{\mathcal{K}}$ as,

$$\mathbf{Adv}_{F_{\mathcal{K}}}(\mathbf{A}) := \left| \Pr_{K \xleftarrow{\$} \mathcal{K}} [\mathbf{A}^{F_K} = 1] - \Pr_{\mathcal{F} \xleftarrow{\$} \mathsf{Func}} [\mathbf{A}^{\mathcal{F}} = 1] \right|.$$

The *maximum prf-advantage of* $F_{\mathcal{K}}$ is defined as

$$\mathbf{Adv}_{F_{\mathcal{K}}}(q, \ell, \sigma) = \max_{\mathbf{A}} \mathbf{Adv}_{F_{\mathcal{K}}}(\mathbf{A})$$

where the maximum is taken over all adversaries $\mathbf{A}$ making at most q queries, of length at most ℓ, and the sum of the lengths of all queries is at most σ. Note that this is an information theoretic definition and we allow an unbounded time adversary. Note that $\mathbf{A}$ is allowed to make adaptive but distinct queries, i.e., for $i \in [2..q]$ it can decide X_i after observing $Y_1, \ldots, Y_{i-1}$ but $X_i \neq X_j$ for $i \neq j \in [q]$. Suppose the q queries are represented by a sequence say $X := (X_1, \ldots, X_q)$, so for any $Y := (Y_1, \ldots, Y_q) \in \mathcal{B}^q$ we have,

$$\Pr_{\mathcal{F} \xleftarrow{\$} \mathsf{Func}} [X \xmapsto{\mathcal{F}} Y] = N^{-q}.$$

2.5 Patarin's Coefficient-H Technique

Let $\mathbf{A}$ be an adversary which makes q distinct queries (possibly adaptive) to F_K. Let the queries be $X_1, \ldots, X_q$ and the corresponding F_K outputs be $Y_1, \ldots, Y_q$. We write $\mathsf{view}(\mathbf{A}^{F_K})$ to denote the q-tuple of pairs $((X_1, Y_1), \ldots, (X_q, Y_q))$ where X_i denotes the i^{th} query and Y_i is the corresponding response.

For any q-tuple of pairs $\tau = ((X_1, Y_1), \ldots, (X_q, Y_q))$, the following probability

$$\mathrm{IP}^{F_K}(\tau) := \Pr_{F_K}[(X_1, \ldots, X_q) \xmapsto{F_K} (Y_1, \ldots, Y_q)]$$

is called the *interpolation probability*, where the probability is taken under the randomness of K. Here we assume that F_K is stateless and so the above probability is independent of the order of the pairs.

Theorem 1 (Coefficient-H Technique)**.** *Let $\mathcal{T}_{good}$ be some set of q-tuples of pairs. Suppose the interpolation probability for a (stateless) oracle $\mathcal{O}$ follows the inequality*

$$\mathrm{IP}^{\mathcal{O}}(\tau) \geq (1 - \epsilon) \cdot \mathrm{IP}^{\mathcal{F}}(\tau) = (1 - \epsilon) N^{-q} \quad \forall\ \tau \in \mathcal{T}_{good}.$$

Then, for any adversary $\mathbf{A}$ we have,

$$\mathbf{Adv}_{F_K}(\mathbf{A}) \leq \epsilon + \Pr[\mathsf{view}(\mathbf{A}^{\mathcal{F}}) \notin \mathcal{T}_{good}].$$

This technique was first introduced by Patarin in his PhD thesis [Pat91] (as mentioned in [Vau03]). The proof of this theorem can be found in [Pat08].

3 CBC-MAC and Its Variants

CBC Function. The CBC (cipher block chaining) function with an oracle $\mathcal{F} \xleftarrow{\$} \mathsf{Func}$, viewed as the key of the construction, takes as input a message $M = (M_1, \ldots, M_b) \in \mathcal{B}^b$ with b blocks and outputs $\mathrm{CBC}^{\mathcal{F},M} := \mathsf{out}_b^{\mathcal{F},M}$. This is inductively computed as follows: $\mathsf{out}_0^{\mathcal{F},M} = 0^n$, and

$$\mathsf{out}_i^{\mathcal{F},M} = \mathcal{F}(\mathsf{in}_i^{\mathcal{F},M}), \quad \mathsf{in}_i^{\mathcal{F},M} = \mathsf{out}_{i-1}^{\mathcal{F},M} \oplus M_i, \quad i \in [b]. \tag{2}$$

Throughout this paper, we call

$$\mathsf{in}^{\mathcal{F},M} = (\mathsf{in}_1^{\mathcal{F},M}, \ldots, \mathsf{in}_b^{\mathcal{F},M}) \text{ and } \mathsf{out}^{\mathcal{F},M} = (\mathsf{out}_1^{\mathcal{F},M}, \ldots, \mathsf{out}_b^{\mathcal{F},M}),$$

the **intermediate input** and **output vectors** respectively, associated to an arbitrary message M and a random function $\mathcal{F}$. *We drop the superscripts $\mathcal{F}$ and M when they are obvious.*

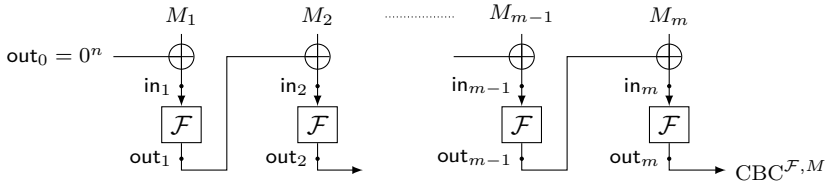

Figure 1: CBC function and its intermediate values where $f \in \mathsf{Func}$.

EMAC Function. The EMAC [BdBB+95] function (E for encrypted) is derived from the CBC function by additionally encrypting the output with another function $\mathcal{F}' \xleftarrow{\$} \mathsf{Func}$ (independent of $\mathcal{F}$). Formally, $\mathrm{EMAC}^{\mathcal{F},\mathcal{F}'}(M) := \mathcal{F}'(\mathrm{CBC}^{\mathcal{F},M})$.

Both CBC and EMAC functions work on inputs from $\mathcal{B}^+$. Black and Rogaway [BR05] suggested three-key variants of EMAC, viz., ECBC, FCBC, and XCBC, to extend the message space to $\sum^*$. Later Kurosawa and Iwata [KI04] gave a two-key variant for EMAC, called TMAC. All these schemes encode a bit sequence $M \in \sum^*$ into a block sequence $\overline{M} := \mathrm{pad}(M)$, defined as

$$\overline{M} = \begin{cases} M\|10^i & \text{if } n \nmid |M| \\ M & \text{otherwise} \end{cases}$$

where i is the smallest non-negative integer such that $M\|10^i \in \mathcal{B}^+$. Let $b = |M|$, $b' = |\overline{M}|_n$, $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ be independently and uniformly chosen from Func, and K, K' be independently and uniformly chosen from $\mathcal{B}$.

ECBC Function. The ECBC [BR00] function is formally defined as,

$$\mathrm{ECBC}^{\mathcal{F}_1,\mathcal{F}_2,\mathcal{F}_3}(M) := \begin{cases} \mathcal{F}_2\big(\mathrm{CBC}^{\mathcal{F}_1,\overline{M}}\big), & n \mid b \\ \mathcal{F}_3\big(\mathrm{CBC}^{\mathcal{F}_1,\overline{M}}\big), & n \nmid b \end{cases}$$

FCBC Function. The FCBC [BR00] function is formally defined as,

$$\mathrm{FCBC}^{\mathcal{F}_1,\mathcal{F}_2,\mathcal{F}_3}(M) := \begin{cases} \mathcal{F}_2\big(\mathrm{CBC}^{\mathcal{F}_1,\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'}\big), & n \mid b \\ \mathcal{F}_3\big(\mathrm{CBC}^{\mathcal{F}_1,\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'}\big), & n \nmid b \end{cases}$$

XCBC Function. The XCBC [BR00] function is formally defined as,

$$\mathrm{XCBC}^{\mathcal{F},K,K'}(M) := \begin{cases} \mathcal{F}\big(\mathrm{CBC}^{\mathcal{F},\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'} \oplus K\big), & n \mid b \\ \mathcal{F}\big(\mathrm{CBC}^{\mathcal{F},\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'} \oplus K'\big), & n \nmid b \end{cases}$$

TMAC Function. The TMAC [KI04] function is formally defined as,

$$\mathrm{TMAC}^{\mathcal{F},K}(M) := \begin{cases} \mathcal{F}\big(\mathrm{CBC}^{\mathcal{F},\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'} \oplus K\big), & n \mid b \\ \mathcal{F}\big(\mathrm{CBC}^{\mathcal{F},\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'} \oplus u \cdot K\big), & n \nmid b \end{cases}$$

where $u \cdot K$ represents the field multiplication of an n-bits non-zero non-one constant u (defined in [KI04]) with K, in a specific representation of $\mathbb{F}_{2^n}$ (specified in [KI04]).

OMAC Function. The OMAC [IK03a] function is formally defined as,

$$\mathrm{OMAC}^{\mathcal{F}}(M) := \begin{cases} \mathcal{F}\big(\mathrm{CBC}^{\mathcal{F},\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'} \oplus \mathcal{F}(0)\big), & n \mid b \\ \mathcal{F}\big(\mathrm{CBC}^{\mathcal{F},\overline{M}_{[b'-1]}} \oplus \overline{M}_{b'} \oplus u \cdot \mathcal{F}(0)\big), & n \nmid b. \end{cases}$$

3.1 PRF Analysis of CBC Variants

In general, the PRF analysis of CBC-MAC variants can be reduced to the analysis of some collision events on the underlying CBC function. This has been the common technique for the PRF analysis of PRP instantiated CBC-MAC variants. We will follow same technique here and establish a relationship between the PRF advantage of CBC-MAC variants and the collision probability of the CBC function. Basically, we show that the PRF advantage of MAC $\in \{\mathrm{EMAC}, \mathrm{ECBC}, \mathrm{FCBC}, \mathrm{XCBC}, \mathrm{TMAC}\}$ is asymptotically tight in the collision probability of the underlying CBC function. First we develop some new terminologies to aid our discussions in rest of the paper. Following that we build the core results of this paper in shape of Lemma 1, 2, Theorem 2, 3 and 4.

Let M_1 and M_2 be two distinct tuple of blocks with block lengths m_1 and m_2 respectively. Let $\mathtt{INcoll}^{\mathcal{F}}(M_1; M_2)$ and $\mathtt{OUTcoll}^{\mathcal{F}}(M_1; M_2)$ denote the events $\mathsf{in}^{\mathcal{F},M_1}_{m_1} = \mathsf{in}^{\mathcal{F},M_2}_{m_2}$ and $\mathsf{out}^{\mathcal{F},M_1}_{m_1} = \mathsf{out}^{\mathcal{F},M_2}_{m_2}$, respectively. We call $\mathtt{INcoll}^{\mathcal{F}}(M_1; M_2)$ the input collision and $\mathtt{OUTcoll}^{\mathcal{F}}(M_1; M_2)$ the output collision events for a pair of messages M_1 and M_2. We similarly define the collision events for a tuple of $q \geq 2$ distinct messages $\mathcal{M} = (M_1, \ldots, M_q)$ as

$$\mathtt{INcoll}^{\mathcal{F}}(\mathcal{M}) = \bigcup_{i \neq j} \mathtt{INcoll}^{\mathcal{F}}(M_i; M_j),$$

and

$$\mathtt{OUTcoll}^{\mathcal{F}}(\mathcal{M}) = \bigcup_{i \neq j} \mathtt{OUTcoll}^{\mathcal{F}}(M_i; M_j).$$

We define **input-collision probability** as $\mathbf{inCP}(\mathcal{M}) = \Pr[\mathtt{INcoll}^{\mathcal{F}}(\mathcal{M})]$ and **output-collision probability** as $\mathbf{outCP}(\mathcal{M}) = \Pr[\mathtt{OUTcoll}^{\mathcal{F}}(\mathcal{M})]$. It is easy to observe that $\mathbf{outCP}(\mathcal{M})$ can be bounded by $\mathbf{inCP}(\mathcal{M})$. When $\mathtt{INcoll}$ is true $\mathtt{OUTcoll}$ is trivially true. Otherwise, $\mathtt{OUTcoll}$ is true when the underlying random function has a collision at the last intermediate output block. More specifically we have,

$$\mathbf{inCP}(\mathcal{M}) \leq \mathbf{outCP}(\mathcal{M}) \leq \mathbf{inCP}(\mathcal{M}) + \frac{q(q-1)}{2N}. \tag{3}$$

Let

$$\mathbf{inCP}_{q,\ell,\sigma} = \max_{\mathcal{M}} \mathbf{inCP}(\mathcal{M})$$

and

$$\mathbf{outCP}_{q,\ell,\sigma} = \max_{\mathcal{M}} \mathbf{outCP}(\mathcal{M})$$

where the maximum is taken over all q-tuple of distinct messages $\mathcal{M}$ having at most ℓ blocks each and the total number of blocks in all the queries is at most σ.

Lemma 1 (PRF–CBC Upper Bound)**.** *For $q, \ell, \sigma \geq 1$ we have,*

1. $\mathbf{Adv}_{\mathrm{EMAC}}(q,\ell,\sigma) \leq \mathbf{inCP}_{q,\ell,\sigma} + \dfrac{q(q-1)}{2N}$.

2. $\mathbf{Adv}_{\mathrm{ECBC}}(q,\ell,\sigma) \leq \mathbf{inCP}_{q,\ell,\sigma} + \dfrac{q(q-1)}{2N}$.

3. $\mathbf{Adv}_{\mathrm{FCBC}}(q,\ell,\sigma) \leq \mathbf{inCP}_{q,\ell,\sigma}$.

4. $\mathbf{Adv}_{\mathrm{XCBC}}(q,\ell,\sigma) \leq \mathbf{inCP}_{q,\ell,\sigma} + \dfrac{q\sigma}{N} + \dfrac{q(q-1)}{2N}$.

5. $\mathbf{Adv}_{\mathrm{TMAC}}(q,\ell,\sigma) \leq \mathbf{inCP}_{q,\ell,\sigma} + \dfrac{q\sigma}{N} + \dfrac{q(q-1)}{2N}$.

3.1.1 Proof of Lemma 1.1, 1.2, and 1.3

Following [BR05, BPR05, Pie06, JN16], we view EMAC, ECBC, and FCBC as instances of the Carter-Wegman (CW) paradigm [WC79]. We discuss the implication of this for each of the schemes below:

1. **EMAC and ECBC:** In case of EMAC and ECBC, we consider the output of the underlying CBC function as the output of an almost-universal hash function. For EMAC the final random function $\mathcal{F}'$ acts on the output of this hash function.

   For ECBC the final random function is either $\mathcal{F}_2$ or $\mathcal{F}_3$ depending upon the padding result. Also observe that in this case the output collision on the CBC outputs for two messages, one each from $\mathcal{B}^+$ and $\sum^*$ is of no use for the adversary as the final outputs are independent due to the independence of $\mathcal{F}_2$ and $\mathcal{F}_3$. So we assume that the messages are from $\mathcal{B}^+$. Now using CW paradigm we have,

   $$\mathbf{Adv}_{\mathrm{E}}(q,\ell,\sigma) \leq \mathbf{outCP}_{q,\ell,\sigma}$$

   for $\mathrm{E} \in \{\mathrm{EMAC}, \mathrm{ECBC}\}$. The result follows by simple application of Equation 3.

2. **FCBC:** In case of FCBC, we consider the input to final random function as the output of an almost-universal hash function. Again using similar arguments as in the case of ECBC we assume that the messages are from $\mathcal{B}^+$. For the ith and jth messages (M_i, M_j) we have to bound the probability of

   $$\mathrm{CBC}^{\mathcal{F}_1, M_{i,[m_i-1]}} \oplus M_{i,m_i} = \mathrm{CBC}^{\mathcal{F}_1, M_{j,[m_j-1]}} \oplus M_{j,m_j}.$$

   This is nothing but

   $$\begin{aligned} \mathsf{out}_{m_i-1}^{M_i} \oplus M_{i,m_i} &= \mathsf{out}_{m_j-1}^{M_j} \oplus M_{j,m_j} \\ \mathsf{in}_{m_i}^{M_i} &= \mathsf{in}_{m_j}^{M_j} \end{aligned}$$

   Here we dropped the superscript $\mathcal{F}_1$ as it is obvious. Now using CW paradigm we have,

   $$\mathbf{Adv}_{\mathrm{FCBC}}(q,\ell,\sigma) \leq \mathbf{inCP}_{q,\ell,\sigma}.$$

3.1.2 Proof of Lemma 1.4 and 1.5

We prove the result for XCBC here. The proof for TMAC can be obtained by slight modification of this proof. Let $\mathcal{T}_{good} := \{(M_1, T_1), \ldots, (M_q, T_q)\}$ be the set of all input output pairs with input messages $\mathcal{M} = (M_1, \ldots, M_q) \in (\sum^+)^q$ and outputs $T = (T_1, \ldots, T_q) \in \mathcal{B}^q$ such that M_i's are distinct and T_i's are also distinct. Also let $\overline{\mathcal{M}} := (\overline{M}_1, \ldots, \overline{M}_q) \in (\mathcal{B}^+)^q$ be the padded version of $\mathcal{M}$, where for all $i \in [q]$, $|\overline{M}_i| = m_i \leq \ell$, and $\sum_{i=1}^q m_i \leq \sigma$. Trivially, random function $\mathcal{F}$ returns a collision pair on any q distinct queries with probability at most $\binom{q}{2}2^{-n}$ for any adversary $\mathbf{A}$. Thus,

$$\Pr[\text{view}(\mathbf{A}^{\mathcal{F}}) \notin \mathcal{T}_{good}] \leq \frac{q(q-1)}{2^{n+1}}.$$

Using coefficient H-technique, now we only need to bound the relationship between the interpolation probabilities. We fix $\tau := (\mathcal{M}, T) \in \mathcal{T}_{good}$. Let

$$\overline{\mathsf{in}} = \left(\mathsf{in}_{m_1}^{\overline{M}_1} \oplus K_1, \ldots, \mathsf{in}_{m_q}^{\overline{M}_q} \oplus K_q\right),$$

where

$$K_i = \begin{cases} K & \text{if } M_i \in \mathcal{B}^+ \\ K' & \text{otherwise.} \end{cases}$$

Let $\texttt{Fresh}$ denote the event,

$$\forall\ r, r' \in [q] \text{ and } i \in [m_{r'}],\ \left(\overline{\mathsf{in}}_r \neq \mathsf{in}_i^{\overline{M}_{r'}}\right)\ \wedge\ \left(r \neq r' \implies \overline{\mathsf{in}}_r \neq \overline{\mathsf{in}}_{r'}\right).$$

Clearly the output of XCBC is completely random, when $\texttt{Fresh}$ holds. Now it is easy to see that,

$$\Pr_{\mathcal{F},K,K'}[\mathcal{M} \overset{\text{XCBC}}{\longmapsto} T \mid \texttt{Fresh}] = \Pr_{\mathcal{F},K,K'}[\overline{\mathsf{in}} \overset{\mathcal{F}}{\longmapsto} T \mid \texttt{Fresh}] = \frac{1}{N^q}.$$

So the interpolation probability of XCBC can be written as,

$$\begin{aligned} \Pr_{\mathcal{F},K,K'}[\mathcal{M} \overset{\text{XCBC}}{\longmapsto} T] &= \Pr_{\mathcal{F},K,K'}[\mathcal{M} \overset{\text{XCBC}}{\longmapsto} T \mid \texttt{Fresh}] \times \Pr_{\mathcal{F},K,K'}[\texttt{Fresh}] \\ &= \frac{1 - \Pr_{\mathcal{F},K,K'}[\neg\texttt{Fresh}]}{N^q}. \end{aligned}$$

Once we upper bound $\Pr_{\mathcal{F},K,K'}[\neg\texttt{Fresh}]$, we are done by the application of coefficient-H technique. Now observe that the event $\neg\texttt{Fresh}$ holds if one of the following three events occur:

1. $\mathtt{B}_1 := \exists\ r, r' \in [q]$, such that $r \neq r'$, $M_r, M_{r'} \in \mathcal{B}^+$ (or $M_r, M_{r'} \notin \mathcal{B}^+$), and $\overline{\mathsf{in}}_r = \overline{\mathsf{in}}_{r'}$.
2. $\mathtt{B}_2 := \exists\ r, r' \in [q]$, such that $r \neq r'$, $M_r \in \mathcal{B}^+$ and $M_{r'} \notin \mathcal{B}^+$, and $\overline{\mathsf{in}}_r = \overline{\mathsf{in}}_{r'}$.
3. $\mathtt{B}_3 := \exists\ r, r' \in [q]$ and $i \in [m_{r'} - 1]$, such that $\overline{\mathsf{in}}_r = \mathsf{in}_i^{\overline{M}_{r'}}$.

$\Pr[\mathtt{B}_1]$: Note that for $\mathtt{B}_1$ the masking keys do not play any role, and there is actually a collision on the output of the underlying CBC function. So we can bound $\Pr[\mathtt{B}_1]$ in terms of **inCP**, i.e,

$$\Pr[\mathtt{B}_1] \leq \mathbf{inCP}_{q,\ell,\sigma}.$$

$\Pr[\mathtt{B}_2]$: For a fix $f \in \mathsf{Func}$, the conditional probability of $\mathtt{B}_2$ is dependent only on the randomness of the masking keys. Further the two masking keys K and K' are uniformly and independently distributed over $\mathcal{B}$. Now it is easy to see that

$$\Pr[\mathtt{B}_2 \mid \mathcal{F} = f] \leq \sum_{r<r' \in [q]} \Pr\left[K \oplus K' = \mathsf{in}_{m_r}^{\overline{M}_r} \oplus \mathsf{in}_{m_{r'}}^{\overline{M}_{r'}} \mid \mathcal{F} = f\right] = \frac{\binom{q}{2}}{N}.$$

Summing over all functions f,

$$\Pr[\mathtt{B}_2] \quad = \sum_{f \in \mathsf{Func}} \Pr_{\mathcal{F},K,K'}[\mathtt{B}_2 \mid \mathcal{F} = f] \times \Pr[\mathcal{F} = f] \quad \leq \frac{\binom{q}{2}}{N}.$$

$\Pr[\mathtt{B}_3]$: For a fix $f \in \mathsf{Func}$, the conditional probability of $\mathtt{B}_3$ is dependent only on the randomness of the masking keys (either one of them). Now it is easy to see that

$$\Pr[\mathtt{B}_3 \mid \mathcal{F} = f] \leq \sum_{r \in [q]} \sum_{\substack{r' \in [q] \\ i \in [m_{r'}-1]}} \Pr_{K_r}\left[K_r = \mathsf{in}_{m_r}^{\overline{M}_r} \oplus \mathsf{in}_i^{\overline{M}_{r'}} \mid \mathcal{F} = f\right] \leq \frac{q\sigma}{N}.$$

Summing over all functions f,

$$\Pr[\mathtt{B}_3] \quad = \sum_{f \in \mathsf{Func}} \Pr_{\mathcal{F},K,K'}[\mathtt{B}_3 \mid \mathcal{F} = f] \times \Pr[\mathcal{F} = f] \quad \leq \frac{q\sigma}{N}.$$

So we have,

$$\Pr[\neg \mathtt{Fresh}] \leq \mathbf{inCP}_{q,\ell,\sigma} + \frac{q\sigma}{N} + \frac{\binom{q}{2}}{N}.$$

■

Remark 1. Note that the PRF analysis of OMAC is missing in lemma 1. Our proof technique cannot be applied directly in case of OMAC. We bound the probability of getting a collision at the input block of the final function. For CBC-MAC variants (other than OMAC) this can be argued using the randomness of the independent random functions or the auxiliary keys. In case of OMAC, $\mathcal{F}(0)$ is used to mask the final internal input block. Whenever the first message block is 0, $\mathcal{F}(0)$ is already defined, hence the current proof technique will not work. Having said that, we believe that identical upper bound should hold for OMAC also.

Lemma 2 (PRF–CBC Lower Bound)**.** *Let $q, \ell \geq 1$ and $\mathcal{M} := (M_1, \ldots, M_q)$ be a q-tuple of distinct messages such that for $i \in [q]$, $M_i \in \mathcal{B} \times (0^n)^{\ell-1}$. Then $\forall$ MAC $\in$ $\{$EMAC, ECBC, FCBC, XCBC, TMAC, OMAC$\}$, we have*

$$\mathbf{Adv}_{\mathrm{MAC}}(q,\ell) \geq \mathbf{inCP}(\mathcal{M})\left(1 - \frac{q(q-1)}{2N}\right).$$

Proof. To show the lower bound we present an adversary $\mathbf{A}$ that attains the claimed PRF advantage using the given message tuple. Consider the following attack algorithm for MAC:

1. $\mathbf{A}$ queries $M_i \in \mathcal{M}$ and observes the corresponding output T_i.
2. If $T_i = T_j$ for some $j < i$ then $\mathbf{A}$ returns 1.

Let $\mathbf{p}_{\mathrm{MAC}}$ and $\mathbf{p}_{\mathcal{F}}$ denote $\Pr[\mathbf{A}^{\mathrm{MAC}} = 1]$ and $\Pr[\mathbf{A}^{\mathcal{F}} = 1]$ respectively. Then we know that,

$$\mathbf{Adv}_{\mathrm{MAC}}(q,\ell,\sigma) \geq \left|\mathbf{p}_{\mathrm{MAC}} - \mathbf{p}_{\mathcal{F}}\right|.$$

For MAC $\in \{\mathrm{EMAC}, \mathrm{ECBC}, \mathrm{FCBC}\}$ we have,

$$\begin{aligned}
&\left|\mathbf{p}_{\mathrm{MAC}} - \mathbf{p}_{\mathcal{F}}\right| \\
&\qquad = \left|\mathbf{outCP}(\mathcal{M}) + (1 - \mathbf{outCP}(\mathcal{M})) \cdot \mathbf{cp}_q - \mathbf{cp}_q\right| && (4)\\
&\qquad = \left|\mathbf{outCP}(\mathcal{M}) \cdot (1 - \mathbf{cp}_q)\right| && (5)\\
&\qquad \geq \left|\mathbf{inCP}(\mathcal{M}) \cdot (1 - \mathbf{cp}_q)\right| && (6)\\
&\qquad \geq \mathbf{inCP}(\mathcal{M}) \cdot \left(1 - \frac{q(q-1)}{2N}\right) && (7)
\end{aligned}$$

Here we use equation (3) from (5) to (6) and equation (1) from (6) to (7). We can have similar analysis for MAC $\in \{\text{XCBC}, \text{TMAC}, \text{OMAC}\}$ by replacing **outCP** with **inCP** in (4). ■

Note that due to the choice of messages (a non-zero block followed by $\ell - 1$ zero blocks) in the attack, the CBC function can be viewed as an iterated random function $f^{(\ell)}$. In other words, our attack also applies on the general iterated random function. Lemma 1 and 2 show that the PRF advantages of EMAC, ECBC, FCBC, XCBC, and TMAC are tight in **inCP** of CBC function.

3.2 Main Results of This Paper

Now we state the main technical results of this paper. The following theorems quantify $\mathbf{inCP}_{q,\ell,\sigma}$. The proofs are postponed to later sections.

Theorem 2 (Upper Bound Theorem). *Let $q, \ell, \sigma \geq 1$. Let $\mathcal{M} = (M_1, \ldots, M_q)$ be a q-tuple of distinct messages such that $M_i \in \mathcal{B}^{m_i}$, $1 \leq m_i \leq \ell$ for all $i \in [q]$, and $\sum_{i=1}^{q} m_i \leq \sigma$. Then we have,*

$$\mathbf{inCP}_{q,\ell,\sigma}(\mathcal{M}) \leq \frac{q\sigma}{N} + \frac{\ell\sigma}{N} + \frac{8q\ell^3\sigma}{N^2}.$$

Theorem 3 (Lower Bound Theorem). *Let $q, \ell, \sigma \geq 1$. Let $\mathcal{M} = (M_1, \ldots, M_q)$ be a q-tuple of distinct messages such that $M_i \in \mathcal{B} \times (0^n)^{\ell-1}$. Then we have,*

$$\begin{aligned}\mathbf{inCP}(\mathcal{M}) \geq \binom{q}{2}\frac{\ell-1}{N}\exp\Big(-\frac{4\ell^2}{N}\Big) - 3\binom{q}{3}\left(\frac{2\ell^2}{N^2} + \frac{6\ell^6}{N^3}\right)\\ - \frac{1}{2}\binom{q}{2}\binom{q-2}{2}\left(\frac{\ell^2}{N^2} + \frac{6\ell^3 + 2\ell^5}{N^3} + \frac{28\ell^8}{N^4}\right)\end{aligned}$$

For $\ell, q \geq 3$, $\frac{q^2\ell}{N} < 1$ and $\ell < \min(\frac{N}{5184}, \frac{N^{\frac{1}{2}}}{4\sqrt{3}}, \frac{N^{\frac{1}{3}}}{\sqrt[3]{36}})$, the above expression is at least $\frac{q^2\ell}{12N}$. Further if we take $\ell = \frac{\sigma}{q}$, the expression is at least $\frac{q\sigma}{12N}$.

Note that in Lemma 2, the advantage can be lower bounded to $\frac{1}{2}\mathbf{inCP}(\mathcal{M})$ for $q < \sqrt{N}$. Using Lemma 1, 2, and Theorem 2, 3 we have the exact PRF security bounds.

Theorem 4 (PRF Bound). *Let $q, \ell, \sigma \geq 3$, such that $\frac{q^2\ell}{N} < 1$, $q < \sqrt{N}$, $\ell = O(q)$, and $\ell < \min(\frac{N}{5184}, \frac{N^{\frac{1}{2}}}{4\sqrt{3}}, \frac{N^{\frac{1}{3}}}{\sqrt[3]{36}}, q)$. Then $\forall$ MAC $\in \{\text{EMAC}, \text{ECBC}, \text{FCBC}, \text{XCBC}, \text{TMAC}\}$, the PRF advantage of* MAC *is asymptotically tight in terms of q, ℓ and σ, i.e.,*

$$\mathbf{Adv}_{\text{MAC}}(q, \ell, \sigma) = \Theta\Big(\frac{q\sigma}{N}\Big).$$

From the above discussion it is clear that, we are only left with the analysis of the collision probability of CBC. More specifically we have to prove Theorem 2 and 3. In [BPR05] Bellare et al. used graph based counting technique to bound the collision probability. The general idea is to map the collision events for a given input (or output) vector to a graph representing the CBC computation. Later this technique was also employed in [Pie06, JN16]. Following [BPR05, Pie06, JN16] we also use structure graphs to bound the collision probability.

3.2.1 A Note on Preneel and van Oorschot's Claim

Preneel and van Oorschot [PvO99] gave a result on the input collision probability [PvO99, Lemma 2] for generic iterated MACs. Informally, they claim:

Let f be a compression function for $n+m$ to n bits. We view the $(n+m)$-bit input as (h,x) where $|h| = n$ and $|x| = n$. If $f(\cdot, x)$ is a uniform random function for all x, then for $2 \leq q \ll 2^n$ distinct inputs which have the last s blocks in common, the collision probability is approximately $1 - \exp\left(\frac{-q(q-1)s}{2^{n+1}}\right)$.

For a uniform random function $\mathcal{F}$, if we define $f(h,x) = \mathcal{F}(h \oplus x)$, then the resulting iterated MAC is CBC-MAC. So, it would seem that Theorem 2 and 3 follow from the above claim (by using $s = O(\ell)$).

However, we observe that the proof of [PvO99, Lemma 2] is flawed. In particular, consider the case, $s > 0$. An output collision implies collision at any of the last $s+1$ iterations. Now, the authors incorrectly argue that (non-)collision event at any of these indices are independent of others. This argument only works if we can ensure that the internal inputs are all distinct for any message. Indeed, it is possible that the inputs at the concerned index collide with some previous inputs, which may lead to a trivial collision at the concerned index. So, the argument is not correct in general.

Consider the CBC-MAC case. The compression function f is based on a length-preserving random function $\mathcal{F}$ with n-bit input. Here, we map $2n$-bit to n-bit, which clearly means that there is a non-zero probability that two internal inputs collide for a given message.

Now, suppose f is based on a compressing random function $\mathcal{F}$ from $n+m$ to n bits, where $m \geq n$. In this case, Preneel and van Oorschot's argument works if we domain separate the input at each index by using counter-based encoding. In fact, the proof is straightforward in this case.

4 Structure Graph

In this section we introduce and setup a tool from [BPR05, Pie06, JN16], called structure graph, that will aid our analysis of the CBC collision probability.

4.0.1 A Note on Directed Edge-Labeled Graph

A directed edge-labeled graph is a pair $G := (\mathcal{V}, \mathcal{E})$ with $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V} \times \mathcal{L}$ where $\mathcal{V}$ is the set of vertices, $\mathcal{L}$ is the set of edge labels, and $\mathcal{E}$ is the set of edges along with their corresponding labels. In this paper we will consider only those directed edge-labeled graphs where for every vertices $u, v \in V$ there exists at most one label $a \in \mathcal{L}$ with $(u,v,a) \in \mathcal{E}$. We also write $u \xrightarrow{a} v$ to mean that $(u,v,a) \in \mathcal{E}$.

Convention: *By abusing notation, $\mathcal{E}$ also denotes the set of unlabeled edges and the label a of the edge $e := (u,v)$ is expressed as $L_G(e)$ (this notation makes sense as there is a unique choice of the label for an edge) or simply $L(e)$ whenever the graph is understood.*

For an edge $e := (u,v)$, vertex u (or v) is called a *predecessor* (or *successor*) of v (or u respectively). An edge (u,v) is called a *loop* if $u = v$. We define two sets:

1. Predecessor set of a vertex v is $\mathsf{nbd}(* \to v) := \{u : (u,v) \in \mathcal{E}\}$.
2. Similarly we define $\mathsf{nbd}(v \to *) := \{u : (v,u) \in \mathcal{E}\}$, the successor set of v.

Definition 3 (Degree)**.** In-degree $\mathbf{deg}_{\text{in}}(v)$ of a vertex v is defined as the size of $|\mathsf{nbd}(* \to v)|$, i.e., $\mathbf{deg}_{\text{in}}(v) = |\mathsf{nbd}(* \to v)|$. Similarly, out-degree $\mathbf{deg}_{\text{out}}(v)$ is defined as $\mathbf{deg}_{\text{out}}(v) = |\mathsf{nbd}(v \to *)|$. Degree $\mathbf{deg}$ of v is defined as the sum of $\mathbf{deg}_{\text{in}}(v)$ and $\mathbf{deg}_{\text{out}}(v)$, i.e., $\mathbf{deg} = \mathbf{deg}_{\text{in}}(v) + \mathbf{deg}_{\text{out}}(v)$.

Definition 4 (Walk, Path and Cycle)**.** A *walk of length* s is defined as a vertex sequence $w := (w_0, \ldots, w_s)$, such that $w_{i-1} \to w_i$ for all $i \in [s]$. We define label of the walk as $L(w) := (a_1, \ldots, a_s)$ where $a_i = L(w_{i-1}, w_i)$, $i \in [s]$. A *subwalk* is defined as a consecutive subsequence $w_{[a..b]}$ of a walk $w_{[0..s]}$ where $0 \le a \le b \le s$. A walk sequence $w_{[0..s]}$ is said to be: a *path* if it has $s+1$ distinct nodes; a *cycle* if it has s distinct nodes and $w_s = w_0$. Note that a loop is also a cycle with $s = 1$.

Definition 5 (t-unicycle)**.** A t-unicycle is a connected directed graph $G = (\mathcal{V}, \mathcal{E})$ where $\forall\ v \in \mathcal{V},\ \mathbf{deg}_{\text{out}}(v) \le 1\ \vee\ \mathbf{deg}_{\text{in}}(v) = 0$, and $\mathcal{E}$ is a union of cycle C and t distinct paths $P_1, \cdots, P_t$ where exactly one endpoint of P_i is a vertex of C and the other endpoint must have zero in-degree. The paths P_i may not be disjoint.

Definition 6 (Isomorphism)**.** Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $G_2 = (\mathcal{V}_1^*, \mathcal{E}_2)$ be two directed graphs. A function $\alpha : \mathcal{V}_1 \to \mathcal{V}_1^*$ is an isomorphism from G_1 to G_2 if α is a bijection and $(u, v, a) \in \mathcal{E}_1$ if and only if $(\alpha(u), \alpha(v), a) \in \mathcal{E}_2$. In this case, we write $G_1 \cong G_2$.

When α is an injective function we can restrict the range set of α such that the restricted range set is the image set of α. This makes the function bijective. We call the graph G_2' so obtained as **α-transformed** G_1 and we write $G_2' = \alpha(G_1)$.

Definition 7 (Function Graph)**.** A directed edge-labeled graph $G = (\mathcal{V}, \mathcal{E})$ is called a *function graph* if

$$\forall\ u \in \mathcal{V},\ \nexists\ v_1, v_2 \in \mathsf{nbd}(u \to *) \text{ such that } v_1 \ne v_2 \text{ and } L_G(u, v_1) = L_G(u, v_2).$$

In other words, for every vertex u and a label a we can find at most one successor v for which the label of the edge (u, v) is a.

This observation can be extended over walks in a function graph G as follows:

$$w_{1,0} = w_{2,0}, L(w_1) = L(w_2)\ \Rightarrow\ \ w_1 = w_2.$$

So if there is a walk with label M then it must be unique and we call such a walk an M-walk (Note, here M is a sequence of message blocks).

4.1 Input-Structure Graph

Notations in this section. Let $\mathcal{M} = (M_1, \ldots, M_q)$ be a q-tuple of distinct messages such that $M_i \in \mathcal{B}^{m_i}$, $1 \le m_i \le \ell$ for all $i \in [q]$, and $\sum_{i=1}^{q} m_i \le \sigma$.

4.1.1 Intermediate Inputs and Outputs

We extend the definition of the CBC function (equation (2), where $\mathcal{F} \in \mathsf{Func}$) for any message $M_i \in \mathcal{M}$ over any $f \in \mathsf{Func}_\perp$ by setting $\mathsf{in}_0^{f,M_i} = \perp$, and then following the normal CBC computation. Recall that this extension does not hamper our analysis, as the uniform distribution over $\mathsf{Func}_\perp$ follows the same p.m.f as Func. This extension may (or may not) look artificial at the moment, but it will greatly simplify some of the definitions and proofs that we discuss later. From now onwards, f and $\mathcal{F}$ will have their usual meaning, but over the set $\mathsf{Func}_\perp$.

Index Set. We define the *index set*

$$\mathcal{I} = \{(r, i) : r \in [q], i \in [0..m_r]\}$$

and the dictionary order $\prec$ on it as follows: $(r, i) \prec (r', i')$ if $r < r'$ or $r = r'$ and $i < i'$. Let X be a sequence over this index set. For any $r \in [q]$, we denote the subsequence $(X_{r,0}, \ldots, X_{r,m_r})$ by $X_{r,*}$.

Sequences for Intermediate Inputs and Outputs. We denote the *sequence of intermediate outputs* and *inputs* over the index set $\mathcal{I}$ as $\mathsf{out}^{f,\mathcal{M}}$ and $\mathsf{in}^{f,\mathcal{M}}$ respectively where

$$\mathsf{out}^{f,\mathcal{M}}_{r,*} = \mathsf{out}^{f,M_r}, \quad \mathsf{in}^{f,\mathcal{M}}_{r,*} = \mathsf{in}^{f,M_r}, \quad \forall r \in [q].$$

For a tuple of messages $\mathcal{M}$, a block sequence X is said to be **realizable** if there exists a function $f \in \mathsf{Func}_\perp$, such that $X = \mathrm{in}^{f,\mathcal{M}}$. Clearly, for all $r \in [q]$, $X_{r,0} = \perp$ for a realizable X. We denote the set of all realizable block sequences for $\mathcal{M}$ by $\mathsf{IN}^{\mathcal{M}}$. The equivalence relation $\sim_X$ induced by a realizable sequence X is called the **collision relation**. Later we try to bound the number of realizable sequences mapping to a collision relation.

4.1.2 (Block-Vertex) Input-Structure Graphs

Definition 8 (Block-Vertex Input-Structure Graph)**.** The **block-vertex input-structure** (BINS) **graph** $\mathsf{BINstruct}^{f,\mathcal{M}}$ for a function f and a message tuple $\mathcal{M}$ is defined by the set of labeled edges

$$\mathcal{E} := \cup_{r=1}^{q} \{(\mathrm{in}_{r,i-1}, \mathrm{in}_{r,i}, M_{r,i}) : i \in [m_r]\}.$$

A BINS graph is a graph theoretic representation of the intermediate input vector $\mathsf{in}^{f,\mathcal{M}}$. Clearly, $\perp \in \mathcal{V}$ has zero in-degree and positive out-degree, and $\mathsf{BINstruct}^{f,\mathcal{M}}$ is a union of M_i-walks, for $i \in [q]$. Note that as explained below,

$$u \xrightarrow{A} v \;\Rightarrow\; f(u) \oplus A = v. \tag{8}$$

So, *for every $u \in V$, all outward edges (similarly for inward edges) have distinct edge labels.* Using this property, it is easy to see that the BINS **graph is a function graph**. So **the walks are unique** and we denote them by w_{M_i} or simply w_i whenever the message tuple is understood. See Fig. 2 for a single message (i.e., $q = 1$) in which the input vector is stored in a directed graph. We denote the set of all BINS graphs by $\mathsf{BINstruct}^{\mathcal{M}}$. While

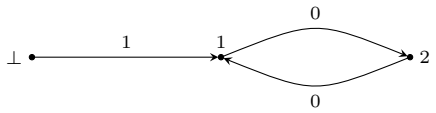

Figure 2: Let $M_1 = (1, 0, 0, 0, 0)$ and $f(1) = 2$, $f(2) = 1$. For any such f, we have $\mathsf{out}^f = (0, 2, 1, 2, 1, 2)$ and $\mathsf{in}^f = (\perp, 1, 2, 1, 2, 1)$. However, the graph consists of three vertices $\{\perp, 1, 2\}$ and edge set $\mathcal{E} = \{(\perp, 1, 1), (1, 2, 0), (2, 1, 0)\}$.

storing the intermediate input sequence as a set of labeled edges, we may loose the order as well as the repetition of the elements. Interestingly, we see that we can uniquely reconstruct the intermediate input sequence from such an edge-labeled graph by using uniqueness of M_i-walks. More precisely, $\mathsf{in}^f_{r,i} = w_{r,i}$. Further the collision relation induced by the input sequence can also be rebuilt by using the M_i-walks. The following lemma is a simple yet important result.

Proposition 2. *Let X and Y be two realizable sequences and G and G' be the associated* BINS *graphs. Then,*

$$\sim_X = \sim_Y \iff G \cong G'.$$

Proof. The proof is left as an exercise for the reader. ■

Definition 9 (Input-Structure Graph)**.** For every vertex v of a BINS graph $G = (\mathcal{V}, \mathcal{E})$, we define a mapping $\alpha : \mathcal{V} \to \mathcal{I}$ as $\alpha_v = \alpha(v) = (r, i)$ where (r, i) is the minimum index such that $w_{r,i} = v$. Clearly, it is an injective mapping with an image set say $\mathcal{V}^*$. Input-structure (INS) graph $G^* = (\mathcal{V}^*, \mathcal{E}^*)$ associated to a BINS graph G is the α-transformed G, i.e., $G^* = \alpha(G)$. The INS graph associated with f and $\mathcal{M}$ is denoted by $\mathsf{INstruct}^{f,\mathcal{M}}$.

Example 1. Let $M_1 = (M_{1,1}, M_{1,2}, M_{1,3}, M_{1,2}, M_{1,4})$ and $M_2 = (M_{2,1})$ be two messages such that $M_{1,1}, M_{1,2}, M_{1,3}, M_{1,2}, M_{1,4}$ and $M_{2,1}$ are distinct. For $f \in \mathsf{Func}_\perp$ let $\mathsf{in}^{f,M_1} = (Y_0 = \perp, Y_1, Y_2, Y_1, Y_2, Y_3)$ (Y_1, Y_2, Y_3 are distinct); $\mathsf{out}^{f,M_1} = (0^n, f(Y_1), f(Y_2), f(Y_1), f(Y_2), f(Y_3))$ and $\mathsf{in}^{f,M_2} = (\perp, Y_3)$;$\mathsf{out}^{f,M_2} = (0^n, f(Y_3))$. The corresponding BINS graph $\mathsf{BINstruct}^f$ is as shown in fig. 3(a) and INS graph $\mathsf{INstruct}^f$ is shown in fig. 3(b).

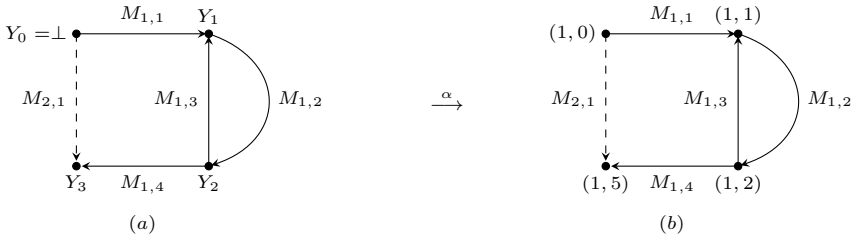

Figure 3: BINS graph $\mathsf{BINstruct}^f$ (a) and the corresponding INS Graph $\mathsf{INstruct}^f$ (b).

Let w_r^* denote the M_r-walk in $\mathsf{INstruct}^{f,\mathcal{M}}$. It is easy to see that an INS graph is again a **union of M_r-walks** w_r^* starting from $(1,0)$.[2] Further, $(1,0)$ (i.e. $\alpha(\perp)$) has zero in-degree. We denote the set of all INS graphs for $\mathcal{M}$ by $\mathsf{INstruct}(\mathcal{M})$. This set has a probability distribution induced due to the uniform distribution on $\mathsf{Func}_\perp$. The following proposition is just an extension of proposition 2.

Proposition 3. *Let X and Y be two realizable sequences such that G and G' are the associated* BINS *graph, and G^* and $G^{*'}$ are the associated* INS *graphs. Then,*

$$\sim_X = \sim_Y \iff G \cong G' \iff G^* = G^{*'}.$$

For a tuple of message $\mathcal{M}$ and a function f, let us try to reconstruct a BINS graph G (there can be more than one candidates for G) from the corresponding INS graph G^*. Specifically we have to find a mapping $Y : \mathcal{V}^* \to \mathcal{B} \cup \{\perp\}$ satisfying certain restrictions. For instance, by definition, $Y_i := Y(i)$ should all be distinct as the valid block label is injective (distinct vertices should get distinct block labels). In addition to this, whenever $e_1 := (u, z), e_2 := (v, z) \in \mathcal{E}$ we must have $f(Y_u) \oplus L(e_1) = f(Y_v) \oplus L(e_2)$. A collision of a graph G^* is defined by such a triple $\delta = (u, v; z)$. The set $\{u, v\}$ is called the *source of the collision* whereas z is called the *head of the collision*. We also say the edges e_1 and e_2 colliding edges. Observe that a collision $\delta = (u, v; z)$ induces a linear restriction $E_\delta : f(Y_u) \oplus f(Y_v) = c_\delta$ on G, where $c_\delta = L(u, z) \oplus L(v, z) \in \mathcal{B}$. By definition of INS graph, this restriction is also preserved in G^*. We denote the set of all collisions of G^* by $\mathcal{C}_{G^*}$, and the set of all linear equations generated by all the collisions by $E(G^*)$, i.e.,

$$E(G^*) := \{E_\delta : \delta \in \mathcal{C}_{G^*}\}.$$

Let $\mathrm{rank}(G^*)$ denote the rank of the system of linear equations in E_{G^*}.

Definition 10 (Accident of an INS graph)**.** We define accident of an INS graph S as $\mathbf{Acc}(S) := \mathrm{rank}(S)$.

Now we mention some important results on INS graphs which will be useful in our analysis ahead. These results have already been proved in [BPR05, Pie06, JN16] for the random permutation case. We prove these results for the random function case in Appendix A.

Lemma 3. *For any* INS *graph S with a accidents,*

$$\Pr[\mathsf{INstruct}^{\mathcal{F}} = S] \le \frac{1}{N^a}.$$

[2]Note that, as per the convention used here and in the preceding discussion $w^*_{r,i} = \alpha(w_{r,i})$.

Lemma 4. *The number of* INS *graphs with a accidents associated to* $\mathcal{M} = (M_1, \ldots, M_t)$ *is at most* $\binom{m}{2}^a$*, where* $\sum_{i=1}^{t} m_i = m$.

Corollary 1. *Let* $a \geq 1$ *be an integer. Then,*

$$\Pr[\mathbf{Acc}(\mathsf{INstruct}^{\mathcal{F}}) \geq a : \mathcal{F} \xleftarrow{\$} \mathsf{Func}] \leq \frac{m^{2a}}{N^a}.$$

5 Proof of Theorem 2 [Upper Bound Theorem]

In this section we upper bound the input-collision probability (and output-collision probability). Recall that the collision relation induced by a realizable sequence is preserved in the corresponding structure graph (using Proposition 3). Further observe that `INcoll` is an example of a specific type of collision relation. So using Proposition 3 we can redefine `INcoll` via structure graphs. For a fixed tuple of message $\mathcal{M}$, `INcoll` is said to be true if there exists some pair of walks w_i and w_j (corresponding to some $M_i, M_j \in \mathcal{M}$) in $\mathsf{INstruct}^{f,\mathcal{M}}$ which share the same last vertex.

For a fixed tuple of messages $\mathcal{M}$, let $\mathsf{INstruct}_a$ denote the set of structure graphs with a accidents. In case of CBC function based on random permutations, different methods [BPR05, Pie06, JN16] were employed to bound the cardinality of $\mathsf{INstruct}_1[\texttt{INcoll}]$[3]. Bellare et al. [BPR05] took a straightforward approach of bounding $|\mathsf{INstruct}_1[\texttt{INcoll}]|$ for two messages and achieved a bound of $d'(\ell)$. In an attempt to get a tighter bound Pietrzak [Pie06] tried to bound the set over groups of messages. Nandi and Jha [JN16] took a much simpler and slight different approach to get tight bounds. We follow the later approach while bounding the collision set.
Let us define the event `Bad` as,

1. for a pair of messages M_i, M_j, $\mathbf{Acc}(\mathsf{INstruct}^{\mathcal{F},(M_i;M_j)}) \geq 2$, or
2. for any message M_i, $\mathbf{Acc}(\mathsf{INstruct}^{\mathcal{F},M_i}) \geq 1$.

We aim to bound the **inCP** in terms of $\Pr[\texttt{Bad}]$ and $\Pr[\neg\texttt{Bad}]$. Specifically we have,

$$\mathbf{inCP}_{q,\ell,\sigma} \leq \Pr_{\mathcal{F}}[\texttt{INcoll}^{\mathcal{F}}(\mathcal{M}) \cap \neg\texttt{Bad}] + \Pr_{\mathcal{F}}[\texttt{Bad}]$$

So we just need to upper bound the following:

1. $\Pr_{\mathcal{F}}[\texttt{Bad}]$. In the first case of `Bad`, we bound the probability by $\sum_{i<j\in[q]} \frac{(m_i+m_j)^4}{N^2}$ (using corollary 1 for all $\binom{q}{2}$ pairs of messages), and in the second case, we bound the probability by $\sum_{i\in[q]} \frac{m_i^2}{N}$(using corollary 1 for all q messages). Finally we have,

$$\begin{aligned}
\Pr_{\mathcal{F}}[\texttt{Bad}] &\leq \sum_{i<j\in[q]} \frac{(m_i+m_j)^4}{N^2} + \sum_{i\in[q]} \frac{m_i^2}{N} \\
&\leq \sum_{i<j\in[q]} \frac{(m_i+m_j)\cdot 8\ell^3}{N^2} + \sum_{i\in[q]} \frac{m_i\cdot\ell}{N} \\
&= \frac{8m(q-1)\ell^3}{N^2} + \frac{m\ell}{N} \\
&\leq \frac{8q\ell^3\sigma}{N^2} + \frac{\ell\sigma}{N}.
\end{aligned}$$

[3] This denotes the set $\{G \in \mathrm{INstruct}_1 : \texttt{INcoll}$ is True for $G\} \subseteq \mathrm{INstruct}_1$

2. $\Pr_{\mathcal{F}}[\texttt{INcoll}^{\mathcal{F}}(\mathcal{M}) \cap \neg\texttt{Bad}]$. $\neg\texttt{Bad}$ implies that $\mathrm{Acc}(\mathsf{INstruct}^{M_i} = w_i) = 0$[4] for $i \in [q]$ or in other words w_i is acyclic. For any pair of messages M_i and M_j, we bound the set $|\mathrm{INstruct}[\texttt{INcoll} \wedge \neg\texttt{Bad}]|$ to at most $\min(m_i, m_j)$ (see the following claim). Since all such graphs must have at most 1 accident, we bound the probability to at most $\frac{\min(m_i,m_j)}{N}$. Hence for q messages we have

$$\Pr_{\mathcal{F}}[\texttt{INcoll}^{\mathcal{F}}(\mathcal{M}) \cap \neg\texttt{Bad}] \leq \sum_{i<j\in[q]} \frac{\min(m_i, m_j)}{N} \leq \frac{q\sigma}{N}.$$

Combining 1 and 2, we have the desired result. ■

Claim: For any pair of messages $M_1, M_2 \in \mathcal{M}$ we have,

$$|\mathrm{INstruct}[\texttt{INcoll} \wedge \neg\texttt{Bad}]| \leq \min(m_1, m_2).$$

Proof. We prove the claim in two cases:

Case 1: $M_1 <_p M_2$. In this case M_1 must be a strict prefix of M_2 as M_1 and M_2 are distinct. Further w_1 is a subwalk of w_2 and we have $w_{1,i} = w_{2,i}$ for $i \in [0..m_1]$. The $\texttt{INcoll}$ event is equivalent to $w_{1,m_1} = w_{2,m_2}$, or $w_{2,m_2} = w_{2,m_1}$. Thus, w_2 must contain a cycle which is not possible. So, $|\mathrm{INstruct}[\texttt{INcoll} \wedge \neg\texttt{Bad}]| = 0$.

Case 2: $M_1 \not<_p M_2$. WLOG assume that $m_1 < m_2$. In this case we must have $p = \mathrm{LCP}(M_1; M_2) < m_1$ otherwise this leads to a cycle in w_2. $\neg\texttt{Bad}$ implies that w_1 and w_2 are paths and $\texttt{INcoll}$ implies that $w_{1,m_1} = w_{2,m_2}$. To get $w_{1,m_1} = w_{2,m_2}$ we must have an accident $(w_{1,i}, w_{2,j}; w_{1,i+1})$ for some $p + 1 \leq i \leq m_1$ and $j = m_2 - m_1 + i$. Therefore, summing over all values of i we have, $|\mathrm{INstruct}(M_1; M_2)[\texttt{INcoll} \wedge \neg\texttt{Bad}]| \leq m_1 \leq \min(m_1, m_2)$.

The result follows from case 1 and 2. ■

6 Proof of Theorem 3 [Lower Bound Theorem]

In this section we show a lower bound of $\frac{q^2\ell}{N}$ on **inCP**. Recall that the attack is also applicable to a general iterated random function. In an independent work Guo et al. [GJMN15] presented a distinguisher for the iterated random function with advantage $\Omega(q^2\ell/N)$. The attack works for $q^2\ell^2 < N$ which makes it ineffective to obtain a significant probability (say $1/2$) for large ℓ. One can repeat the attack to amplify the probability but doing so will lead to a loss in terms of complexity. In Theorem 3 we show that our attack achieves the same bound of $\Omega(q^2\ell/N)$ for much less restrictions on ℓ.

Remark 2. Bellare et al. [BPR05] proved that the CBC-MAC based on random permutation is secure and the advantage is bounded by $O(q^2\ell/N)$ provided $\ell = o(N^{1/3})$. Here we show that there is an attack for CBC based on random function with advantage $\Omega(q^2\ell/N)$. Our idea of the attack algorithm can not be easily extended to CBC based on random permutation. It seems that CBC based on random permutation is more secure than the one based on random function.

[4] From now onwards we use w_i to represent M_i-walk in the INS graph.

6.1 Our Attack Algorithm for CBC collision

Let $\mathcal{M} := (M_1, \ldots, M_q)$ be a q-tuple of messages such that for $i, j \in [q]$ $M_i = (X_i, 0, \cdots, 0) \in \mathcal{B}^\ell$ and $X_i \neq X_j \in \mathcal{B}$. We want to find a lower bound of collision probability that is,

$$\begin{aligned}\mathbf{inCP}(\mathcal{M}) &= \Pr_{\mathcal{F}}[\mathtt{INcoll}^{\mathcal{F}}(\mathcal{M})] = \Pr_{\mathcal{F}}\left[\bigcup_{1\le i<j\le q} \mathtt{INcoll}^{\mathcal{F}}(M_i; M_j)\right] \\ &\geq \sum_{i<j} \Pr_{\mathcal{F}}[\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j)] \\ &- 3\sum_{i<j<k} \Pr_{\mathcal{F}}[\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \mathtt{INcoll}^{\mathcal{F}}(M_j; M_k)] \\ &- \frac{1}{2} \sum_{\substack{i<j,k<m \\ \{i,j\}\cap\{k,m\}=\emptyset}} \Pr_{\mathcal{F}}[\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \mathtt{INcoll}^{\mathcal{F}}(M_k; M_m)] \qquad (9)\end{aligned}$$

where $\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j)$ denotes the event that $\mathsf{in}_\ell^{\mathcal{F},M_i} = \mathsf{in}_\ell^{\mathcal{F},M_j}$.The last inequality follows from Principle of Inclusion-Exclusion and Bonferroni inequality. Now, we need to compute the following bounds,

- Upper bound for $\Pr_{\mathcal{F}}[\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \mathtt{INcoll}^{\mathcal{F}}(M_j; M_k)]$, where i, j and k are distinct.
- Upper bound for $\Pr_{\mathcal{F}}[\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \mathtt{INcoll}^{\mathcal{F}}(M_k; M_m)]$, where i, j, k and m are distinct.
- Lower bound for $\Pr_{\mathcal{F}}[\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j)]$, where i and j are distinct.

We use structure graphs to bound the above mentioned probabilities. Observe that due to our choice of messages we have the following property on $\mathsf{INstruct}^{\mathcal{F},\mathcal{M}}$:

$$\forall v \in \mathcal{V}(\mathsf{INstruct}^{\mathcal{F},\mathcal{M}}) \setminus \{(1,0)\},\ \mathbf{deg}_{\text{out}}(v) \leq 1.$$

This is obvious as all the edge labels (except those involving $(1,0)$) are identical (0^n). Therefore, $\mathsf{INstruct}^{\mathcal{F},\mathcal{M}}$ is either a union of paths or a union of unicycles or both. Note that in either case the graph has no dependent collisions with distinct heads (as that requires $\mathbf{deg}_{\text{out}}(v) \geq 2$ for some v). Further, for a vertex with in-degree $a \geq 1$, we have $a - 1$ independent collisions or accidents.

Remark 3. We summarise some useful properties derived from the above discussion:

1. A union of k paths has at most $k - 1$ collisions. So the number of accidents is $k - 1$.
2. A k-unicycle has exactly k collisions. So the number of accidents is k.
3. A union of a k_1-unicycle and k_2 paths has at most $k_1 + k_2 - 1$ accidents.
4. A union of a k_1-unicycle and a k_2-unicycle has exactly $k_1 + k_2$ accidents.
5. In general k distinct walks (where each vertex v has $\mathbf{deg}_{\text{out}}(v) \leq 1$), can have at most k accidents.

Example 2. In figure 4,

1. S is a 3-unicycle with $\mathbf{Acc}(S) = 3$.
2. S' is a union of 3 paths with $\mathbf{Acc}(S') = 1$.

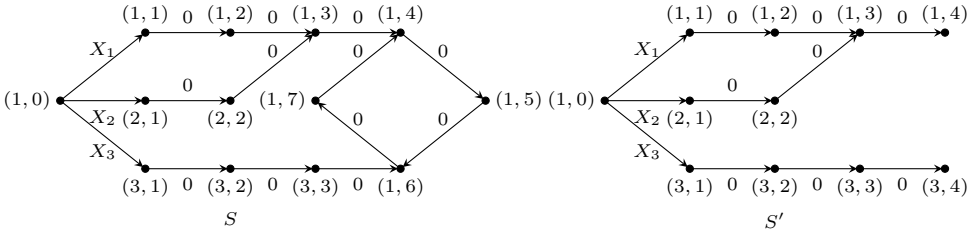

Figure 4: Unicycles and union of paths.

Now we bound the above mentioned probabilities in Lemma 5, 6, and 7. We postpone the proofs of these results to Subsection 6.2.

Lemma 5.

$$\Pr_{\mathcal{F}}[\texttt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \texttt{INcoll}^{\mathcal{F}}(M_j; M_k)] \leq \frac{2\ell^2}{N^2} + \frac{6\ell^6}{N^3}.$$

Lemma 6.

$$\Pr_{\mathcal{F}}[\texttt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \texttt{INcoll}^{\mathcal{F}}(M_k; M_m)] \leq \frac{\ell^2}{N^2} + \frac{6\ell^3 + 2\ell^5}{N^3} + \frac{28\ell^8}{N^4}.$$

Lemma 7.

$$\Pr_{\mathcal{F}}[\texttt{INcoll}^{\mathcal{F}}(M_i; M_j)] \geq \frac{\ell - 1}{N} \exp\Big(-\frac{4\ell^2}{N}\Big).$$

Combining Equation 9 with Lemma 5, 6 and 7 we have,

$$\begin{aligned}\mathbf{inCP}(\mathcal{M}) \geq \binom{q}{2}\frac{\ell-1}{N}\exp\Big(-\frac{4\ell^2}{N}\Big) - 3\binom{q}{3}\left(\frac{2\ell^2}{N^2} + \frac{6\ell^6}{N^3}\right) \\ - \frac{1}{2}\binom{q}{2}\binom{q-2}{2}\left(\frac{\ell^2}{N^2} + \frac{6\ell^3 + 2\ell^5}{N^3} + \frac{28\ell^8}{N^4}\right)\end{aligned}$$

If $\ell, q \geq 3$, $\frac{q^2\ell}{N} < 1$ and $\ell < \min(\frac{N}{5184}, \frac{N^{\frac{1}{2}}}{4\sqrt{3}}, \frac{N^{\frac{1}{3}}}{\sqrt[3]{36}})$, then using the inequality $\exp(-x) \geq 1 - x$ and some algebraic manipulations one can show

$$\mathbf{inCP}(\mathcal{M}) \geq \frac{q^2\ell}{12N}.$$

Further for $\ell = \sigma/q$ this bound becomes $\frac{q\sigma}{12N}$. ■

6.2 Proofs Related to the Lower Bound Theorem

Proof of Lemma 5. Let $\texttt{INcoll}^{i,j,k}$ denote the event $\texttt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \texttt{INcoll}^{\mathcal{F}}(M_j; M_k)$. From remark 3 we know that the number of accidents must be ≤ 3.

1. Observe that $\texttt{INcoll}^{i,j,k}$ requires at least 2 independent collisions (accidents) (in w_i and w_j paths and w_k and $w_i \cup w_j$), so

$$|\mathsf{INstruct}_1[\texttt{INcoll}^{i,j,k}]| = 0.$$

2. Now, it is easy to see that accident 2 graphs are possible only for union of paths, as the number of accidents correspond to the collision between the three paths. There are at most ℓ many choices for collision between w_i and w_j paths, and at most 2ℓ many choices for collision between w_k and $w_i \cup w_j$. This bounds

$$|\mathsf{INstruct}_2[\texttt{INcoll}^{i,j,k}]| \leq 2\ell^2.$$

3. The graph can have 3 accidents iff it is a 3-unicycle. Suppose the cycle is in w_i. Then w_i is determined by the length of cycle and the distance of $w_{i,1}$ from the cycle which gives ℓ^2 choices for w_i. For each such choice we have at most $2\ell^2$ many choices for $w_i \cup w_j$, and at most $3\ell^2$ many choices for $w_i \cup w_j \cup w_k$. This bounds

$$|\mathsf{INstruct}_3[\mathtt{INcoll}^{i,j,k}]| \leq 6\ell^6.$$

The result follows by direct application of Lemma 3. ■

Proof of Lemma 6. Let $\mathtt{INcoll}^{i,j,k,m}$ denote the event $\mathtt{INcoll}^{\mathcal{F}}(M_i; M_j) \wedge \mathtt{INcoll}^{\mathcal{F}}(M_k; M_m)$. From remark 3 we know that the number of accidents must be ≤ 4. We bound the four sets corresponding to the number of accidents as below:

1. Accident 1 graphs are not possible. Hence,

$$|\mathsf{INstruct}_2[\mathtt{INcoll}^{i,j,k,m}]| = 0.$$

2. The accident 2 graphs are possible iff $(w_i \cup w_j) \cap (w_k \cup w_m) = \{(1,0)\}$, where the two accidents correspond to the collision between w_i and w_j, and collision between w_k and w_m. Now there are at most ℓ many choices for collision between w_i and w_j paths, and similarly at most ℓ many choices for collision between w_k and w_m paths. This gives

$$|\mathsf{INstruct}_2[\mathtt{INcoll}^{i,j,k,m}]| \leq \ell^2.$$

3. Accident 3 graphs are possible iff,

   (a) w_i, w_j, w_k, w_m paths collide and the graph is a union of paths. In this case we have at most ℓ many choices for collision between w_i and w_j paths. Similarly we have at most 2ℓ and 3ℓ many choices for collision between w_k and $w_i \cup w_j$, and w_m and $w_i \cup w_j \cup w_m$ respectively. This gives

$$|\mathsf{INstruct}_3[\mathtt{INcoll}^{i,j,k,m}]| \leq 6\ell^3.$$

   (b) $(w_i \cup w_j) \cap (w_k \cup w_m) = \{(1,0)\}$ and $w_i \cup w_j$ is a 2-unicycle and $w_k \cup w_m$ is a union of paths or vice versa. Without loss of generality assume that $w_i \cup w_j$ is a 2-unicycle. Then there exist a cycle in $w_i \cup w_j$. Suppose the cycle is in w_i. Then w_i is determined by the length of cycle and the distance of $w_{i,1}$ from the cycle which gives ℓ^2 choices for w_i. For each such choice we have at most $2\ell^2$ many choices for $w_i \cup w_j$. And there are at most ℓ many choices for collision between w_k and w_m. This gives,

$$|\mathsf{INstruct}_3[\mathtt{INcoll}^{i,j,k,m}]| \leq 2\ell^5.$$

   Combining the two subcases we have,

$$|\mathsf{INstruct}_3[\mathtt{INcoll}^{i,j,k,m}]| \leq 6\ell^3 + 2\ell^5.$$

4. Accident 4 graphs are possible iff,

   (a) $w_i \cup w_j$ and $w_k \cup w_m$ are distinct 2-unicycles. Using similar arguments as used in the previous cases we get a bound of $4\ell^8$.

   (b) w_i, w_j, w_k, w_m form a 4-unicycle. This case can be bounded to $24\ell^8$, using similar approach as used in the previous cases.

   Combining the two subcases we have,

$$|\mathsf{INstruct}_4[\mathtt{INcoll}^{i,j,k,m}]| \leq 28\ell^8.$$

The result follows by direct application of Lemma 3. ■

Proof of Lemma 7. Let $\texttt{INcoll}^{i,j}$ denote the event $\texttt{INcoll}^{\mathcal{F}}(M_i; M_j)$. We are basically interested in the probability that $\mathsf{INstruct}^{\mathcal{F}} \in \mathsf{INstruct}(M_i, M_j)[\texttt{INcoll}]$. Let $\texttt{Acyclic}$ denote the property that some graph $S \in \mathsf{INstruct}(M_i, M_j)[\texttt{INcoll}]$ is acyclic graph, and $\mathsf{INstruct}(M_i, M_j)[\texttt{INcoll} \wedge \texttt{Acyclic}]$ denote the subset of all graphs which satisfy both $\texttt{INcoll}$ and $\texttt{Acyclic}$. Thus, we have

$$\Pr[\mathsf{INstruct}^{\mathcal{F}} \in \mathsf{INstruct}(M_i, M_j)[\texttt{INcoll}]] \geq \Pr[\mathsf{INstruct}^{\mathcal{F}} \in \mathsf{INstruct}(M_i, M_j)[\texttt{INcoll} \wedge \texttt{Acyclic}]]$$

We will lower bound $\Pr[\mathsf{INstruct}^{\mathcal{F}} \in \mathsf{INstruct}(M_i, M_j)[\texttt{INcoll} \wedge \texttt{Acyclic}]]$. First, convince yourself that for all $S \in \mathsf{INstruct}(M_i, M_j)[\texttt{INcoll} \wedge \texttt{Acyclic}]$ we must have $\mathbf{Acc}(S) = 1$ (as $\texttt{INcoll} \wedge \texttt{Acyclic}$ holds and M_i and M_j share a common suffix of length $\ell - 1$).

This accident can happen at any one of the index $2 \leq k \leq \ell$, each contributing exactly one structure graph. Fix an index $2 \leq k \leq \ell$ where the accident occurs and let the corresponding INS graph be S_k. Then, a simple counting shows that the number of valid block labeling for S_k is exactly $(2^n - 2) \dots (2^n - 2k + 2)$. Each such labeling gives a BINS graph G with exactly $2k - 2$ positive out-degree vertices (excluding $\perp$ which is trivial) such that $\alpha(G) = S$. The probability of getting a BINS graph with $2k - 2$ many vertices having positive out-degree is equal to 2^{2n-2kn} (as exactly $2k - 2$ outputs of $\mathcal{F}$ are fixed). Thus, we get

$$\begin{aligned}
\Pr[\mathsf{INstruct}^{\mathcal{F}} \in \mathsf{INstruct}(M_i, M_j)[\texttt{INcoll} \wedge \texttt{Acyclic}]] &= \frac{1}{2^n} \sum_{k=2}^{\ell} \left(1 - \frac{2}{2^n}\right) \cdots \left(1 - \frac{2k-2}{2^n}\right) \\
&\geq \frac{\ell - 1}{2^n} \prod_{k=1}^{2\ell-2} \left(1 - \frac{k}{2^n}\right) \\
&\geq \frac{\ell - 1}{2^n} \left(1 - \frac{2\ell^2}{2^n}\right) \\
&\geq \frac{\ell - 1}{2^n} \exp\left(-\frac{4\ell^2}{2^n}\right),
\end{aligned}$$

where the last inequality follows from $(1 - x) \geq \exp(-2x)$ for $0 < x < 0.5$, and the assumption that $\ell < 2^{\frac{n}{2}-1}$. ■

7 Conclusion and Future Work

As summarized in Table 1 and Table 2, in terms of exact security random permutation vs random function based message authentication codes differ considerably. For EMAC, ECBC and FCBC constructions the random permutation instantiations have optimal security bound $\Theta(\frac{q^2}{2^n})$. The corresponding random function instantiations are considerably less secure, with optimal security bound $\Theta(\frac{q^2}{2^n})$. For XCBC and TMAC constructions random permutation instantiations are at least as secure as random function instantiations. It is an interesting open problem, whether random permutation based upper (or lower) bounds can be improved for XCBC and TMAC.

The upper bounds obtained in this work are of the form $O(\frac{q\sigma}{2^n})$. In this work we have not considered OMAC [IK03a]. For random function based OMAC, our lower bound would hold as it is. It is an open problem whether one can obtain matching upper bound of the form $O(\frac{\ell q^2}{2^n})$ or $O(\frac{q\sigma}{2^n})$.

References

[BdBB+95] A. Berendschot, B. den Boer, J. Boly, A. Bosselaers, J. Brandt, D. Chaum, I. Damgård, M. Dichtl, W. Fumy, M. van der Ham, C. Jansen, P. Landrock, B. Preneel, G. Roelofsen, P. de Rooij, and J Vandewalle. *Final Report of Race Integrity Primitives*, volume 1007 of *Lecture Notes in Computer Science, Springer-Verlag, 1995*. Springer-Verlag, 1995.

[Ber03] Robert Berke. On the security of iterated macs. Diploma thesis, ETH Zurich, 2003.

[BG08] Mihir Bellare and Shafi Goldwasser. Lecture Notes on Cryptography. *Summer course on Cryptography and Computer Security at MIT (1996–2008)*, pages 249–250, 2008.

[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.

[BPR05] Mihir Bellare, Krzysztof Pietrzak, and Phillip Rogaway. Improved Security Analyses for CBC MACs. In *Advances in Cryptology - CRYPTO 2005: 25th Annual International Cryptology Conference, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, pages 527–545, 2005.

[BR00] John Black and Phillip Rogaway. CBC macs for arbitrary-length messages: The three-key constructions. In *Advances in Cryptology - CRYPTO 2000, 20th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2000, Proceedings*, pages 197–215, 2000.

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

[GJMN15] Jian Guo, Jérémy Jean, Nicky Mouha, and Ivica Nikolic. More rounds, less security? *IACR Cryptology ePrint Archive*, 2015:484, 2015.

[IK03a] Tetsu Iwata and Kaoru Kurosawa. OMAC: One-Key CBC MAC. In *Fast Software Encryption, 10th International Workshop, FSE 2003, Lund, Sweden, February 24-26, 2003, Revised Papers*, pages 129–153, 2003.

[IK03b] Tetsu Iwata and Kaoru Kurosawa. Stronger Security Bounds for OMAC, TMAC, and XCBC. In *Progress in Cryptology - INDOCRYPT 2003, 4th International Conference on Cryptology in India, New Delhi, India, December 8-10, 2003, Proceedings*, pages 402–415, 2003.

[ISO11] ISO/IEC. Information Technology – Security Techniques – Message Authentication Codes (MACs) – Part 1: Mechanisms Using A Block Cipher. International Standard ISO/IEC 9797-1:2011, International Organization for Standardization, Geneva, CH, 2011.

[JN16] Ashwin Jha and Mridul Nandi. Revisiting Structure Graphs: Applications to CBC-MAC and EMAC. *J. Mathematical Cryptology*, 10(3–4):157–180, 2016.

[KI03] Kaoru Kurosawa and Tetsu Iwata. Tmac: Two-key cbc mac. In *Cryptographers' Track at the RSA Conference*, pages 33–49. Springer, 2003.

[KI04] Kaoru Kurosawa and Tetsu Iwata. TMAC: Two-Key CBC MAC. *IEICE Transactions*, 87-A(1):46–52, 2004.

[MM07] Kazuhiko Minematsu and Toshiyasu Matsushima. New Bounds for PMAC, TMAC, and XCBC. In *Fast Software Encryption, 14th International Workshop, FSE 2007, Luxembourg, Luxembourg, March 26-28, 2007, Revised Selected Papers*, pages 434–451, 2007.

[Nan09] Mridul Nandi. Improved Security Analysis for OMAC as a Pseudorandom Function. *J. Mathematical Cryptology*, 3(2):133–148, 2009.

[Pat91] Jacques Patarin. *Etude des Générateurs de Permutations Pseudo-aléatoires Basés sur le Schéma du DES*. PhD thesis, Université de Paris, 1991.

[Pat08] Jacques Patarin. The "Coefficients H" Technique. In *Selected Areas in Cryptography, 15th International Workshop, SAC 2008, Sackville, New Brunswick, Canada, August 14-15, Revised Selected Papers*, pages 328–345, 2008.

[Pie06] Krzysztof Pietrzak. A Tight Bound for EMAC. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, pages 168–179, 2006.

[PR00] Erez Petrank and Charles Rackoff. CBC MAC for Real-Time Data Sources. *J. Cryptology*, 13(3):315–338, 2000.

[PvO99] Bart Preneel and Paul C. van Oorschot. On the security of iterated message authentication codes. *IEEE Trans. Information Theory*, 45(1):188–199, 1999.

[Vau03] Serge Vaudenay. Decorrelation: A Theory for Block Cipher Security. *J. Cryptology*, 16(4):249–286, 2003.

[WC79] Mark N. Wegman and Larry Carter. New Classes and Applications of Hash Functions. In *20th Annual Symposium on Foundations of Computer Science, San Juan, Puerto Rico, 29-31 October 1979*, pages 175–182, 1979.

A Proofs of Results on Structure Graph

Lemma (Lemma 3)**.** *For any* INS *graph* S *with* a *accidents,*

$$Pr[\mathsf{INstruct}^{\mathcal{F}} = S] \leq \frac{1}{N^a}.$$

Proof. S is an INS graph with a accidents, i.e., $\mathrm{rank}(S) = a$. We denote the number of vertices in S, excluding $(1, 0)$, with positive out-degree by s. Using linear algebra, we know that some $s - a$ choices of $f(Y_i)$ values will uniquely determine the rest and so the number of valid block labellings is at most N^{s-a}. Any valid choice of Y induces a block-vertex structure graph $G = (\mathcal{V}, \mathcal{E})$ such that $G^* = S$. Note that s is the number of vertices $v \in \mathcal{V}$ with positive out-degree. So exactly N^{N-s} number of functions can result in BINS graph G. Therefore,

$$\mathsf{Pr}[\mathsf{BINstruct}^{\mathcal{F}} = G] = \frac{N^{N-s}}{N^N} = \frac{1}{N^s}. \tag{10}$$

Summing over all BINS graphs G such that the INS graph $\alpha(G) = S$ we have,

$$\mathsf{Pr}[\mathsf{INstruct}^{\mathcal{F}} = S] = \sum_{G:\alpha(G)=S} \mathsf{Pr}[\mathsf{BINstruct}^{\mathcal{F}} = G] \leq \frac{N^{s-a}}{N^s} = \frac{1}{N^a}.$$

■

For an INS graph S we define **traversal** $\mathcal{T}(S)$ as the sequence of vertices $\mathcal{T}(S) := (w^*_{r,i})_{(r,i)\in\mathcal{I}}$. Note that $\mathcal{T}(S)$ implicitly stores the edges: for every $\alpha \in \mathcal{I}$ such that

$\alpha \neq (r, m_r)$ we have $(w^*_\alpha, w^*_{\alpha+1}) \in \mathcal{E}$ with label $M_{\alpha+1}$. We denote the set of edges in $\mathcal{T}(S)$ by $\mathcal{E}(\mathcal{T}(S))$. The sub-sequence $\mathcal{T}_\alpha(S) := \mathcal{T}(S)_\alpha$ of $\mathcal{T}(S)$ is called the partial traversal till $\alpha \in \mathcal{I}$. In $\mathcal{T}(S)$, an output-collision $\delta := (u, v; z)$ can be equivalently written as,

$$\delta = (w^*_i = u, w^*_j = v; w^*_{i+1} = w^*_{j+1} = z),\ i \prec j \in \mathcal{I},$$

where i and j are the smallest such indices. Under this equivalent representation we can define a partial order $\prec_{\mathcal{C}}$ on $\mathcal{C}(S)$ as follows:
For $i, j, i^{'}, j' \in \mathcal{I}$ and $i \prec j$ and $i^{'} \prec j^{'}$, let $\delta = (w^*_i, w^*_j; w^*_{i+1})$ and $\delta' = (w^*_{i'}, w^*_{j'}; w^*_{i'+1})$. $\delta \prec_{\mathcal{C}} \delta'$ if either,

1. $j \prec j^{'}$, or
2. $j = j^{'}$ and $i \prec i^{'}$.

Proposition 4. *Let $S_1, S_2 \in \mathsf{INstruct}^{\mathcal{M}}$ be two* INS *graphs and $\mathcal{T}(S_1)$ and $\mathcal{T}(S_2)$ be their associated traversals. Then,*

$$\forall \alpha \in \mathcal{I} \quad \mathcal{T}_\alpha(S_1) = \mathcal{T}_\alpha(S_2) \iff S_1^{\mathcal{E}_\alpha} = S_2^{\mathcal{E}_\alpha}$$

where $S_i^{\mathcal{E}_\alpha}$ is the edge induced subgraph of S_i with edge set $\mathcal{E}(\mathcal{T}_\alpha(S_i))$. Particularly for $\alpha = (q, m_q)$ we have,

$$\mathcal{T}(S_1) = \mathcal{T}(S_2) \iff S_1 = S_2.$$

Proof. The necessary condition, i.e.,

$$S_1^{\mathcal{E}_\alpha} = S_2^{\mathcal{E}_\alpha} \implies \mathcal{T}_\alpha(S_1) = \mathcal{T}_\alpha(S_2)$$

is trivially true (by definition of traversals). So we focus on the sufficient condition, i.e.,

$$\mathcal{T}_\alpha(S_1) = \mathcal{T}_\alpha(S_2) \implies S_1^{\mathcal{E}_\alpha} = S_2^{\mathcal{E}_\alpha}.$$

Note that $\mathcal{T}_\alpha(S_1) = \mathcal{T}_\alpha(S_2) \implies \mathcal{E}(\mathcal{T}_\alpha(S_1)) = \mathcal{E}(\mathcal{T}_\alpha(S_2))$. As the two edge sets are equal, the edge induced subgraphs must also be equal. Thus $S_1^{\mathcal{E}_\alpha} = S_2^{\mathcal{E}_\alpha}\ \ \forall \alpha \in \mathcal{I}$. ■

Definition 11 (Accident Basis and Dependent Collisions)**.** We define the accident basis $\mathcal{C}^{\mathrm{Acc}}(S)$ of the INS graph S as $C \subseteq \mathcal{C}(S)$ such that $E(C) := \{E_\delta : \delta \in C\}$ is the minimal spanning set of $E(S)$ and the elements of C are smallest with respect to $\prec_{\mathcal{C}}$. Set $\mathcal{D} \subset \mathcal{C}(S)$ is called a set of dependent collisions if $E(\mathcal{D})$ is linearly dependent. Note that $\mathbf{Acc}(S) = |\mathcal{C}^{\mathrm{Acc}}(S)|$ as $E(\mathcal{C}^{\mathrm{Acc}}(S))$ is a basis of $E(S)$.

It is obvious that $\mathcal{C}^{\mathrm{Acc}}(S^{\mathcal{E}_\alpha})$ the accident basis corresponding to the edge induced subgraph $S^{\mathcal{E}_\alpha}$ (equivalently to the partial traversal $\mathcal{T}_\alpha(S)$) is a subset of $\mathcal{C}^{\mathrm{Acc}}(S)$.

Example 3. Let $M_1 = (M_{1,1}, M_{1,2}, M_{1,3}, M_{1,2}, M_{1,4})$ and $M_2 = (M_{2,1})$ be two messages such that $M_{1,1} \oplus M_{1,3} \oplus M_{1,4} \oplus M_{2,1} = 0^n$. For $f \in \mathsf{Func}_\perp$ let S be the INS graph as shown in Figure 5. For S we have,

1. $\mathcal{T}(S) := (w^*_{1,0}, w^*_{1,1}, w^*_{1,2}, w^*_{1,3}, w^*_{1,4}, w^*_{1,5}, w^*_{2,0}, w^*_{2,1})$, where $w^*_{1,0} = (1, 0),\ w^*_{1,1} = (1, 1),\ w^*_{1,2} = (1, 2),\ w^*_{1,3} = w^*_{1,1},\ w^*_{1,4} = w^*_{1,2},\ w^*_{1,5} = (1, 5),\ w^*_{2,0} = w^*_{1,0},\ w^*_{2,1} = w^*_{1,5}$.
2. $\mathcal{C}(S) = \big\{((1, 0), (1, 2); (1, 1)), ((1, 2), (1, 0); (1, 5))\big\}$. The equivalent representation in $\mathcal{T}(S)$ is

$$\big\{(w^*_{1,0}, w^*_{1,2}; w^*_{1,1} = w^*_{1,3}),\\ (w^*_{1,4}, w^*_{2,0}; w^*_{1,5} = w^*_{2,1})\big\}.$$

3. $E(S) = \{f(Y_{1,0}) \oplus f(Y_{1,2}) = M_{1,1} \oplus M_{1,3}; f(Y_{1,0}) \oplus f(Y_{1,2}) = M_{1,4} \oplus M_{2,1}\}$.
4. Clearly, $\mathbf{Acc}(S) = \mathrm{rank}(S) = 1$.
5. $\mathcal{C}^{\mathrm{Acc}}(S) = \{((1,0),(1,2);(1,1))\}$.

.

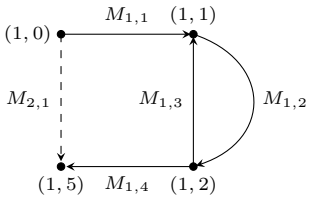

Figure 5: The INS graph, S.

Lemma (Lemma 4)**.** *The number of* INS *graphs with a accidents associated to* $\mathcal{M} = (M^1, \ldots, M^t)$ *is at most* $\binom{m}{2}^a$*, where* $\sum_{i=1}^{t} m_i = m$.

Proof. The proof becomes trivial once we show that each structure graph has a unique accident basis and distinct graphs have distinct accident basis. In other words, we need to show that the mapping from the set of structure graphs to the set of accident basis is injective. It is easy to see that each structure graph has a unique accident basis (by the definition of accident basis). We know show that distinct structure graphs have distinct accident basis.

Claim: Let S_1 and S_2 be two structure graphs. Then,

$$\mathcal{C}^{\mathrm{Acc}}(S_1) = \mathcal{C}^{\mathrm{Acc}}(S_2) \implies S_1 = S_2.$$

Using Proposition 4 it is sufficient to show that

$$\mathcal{C}^{\mathrm{Acc}}(S_1) = \mathcal{C}^{\mathrm{Acc}}(S_2) \implies \mathcal{T}(S_1) = \mathcal{T}(S_2).$$

We prove the claim by induction on the dictionary order over the index set $\mathcal{I}$. Let $\alpha \in \mathcal{I}$. Suppose $\mathcal{T}_\beta(S_1) = \mathcal{T}_\beta(S_2) \ \ \forall \beta \prec \alpha$. If $\alpha = (r, m_r)$ for some $r \in [q]$, then the next vertex on $\mathcal{T}(S_1)$ i.e. $w_\alpha^{*1} = (1,0) = w_\alpha^{*2}$, the next vertex on $\mathcal{T}(S_2)$. Thus, $\mathcal{T}_\alpha(S_1) = \mathcal{T}_\alpha(S_2)$. Suppose $\alpha = (r,i)$ for some $r \in [q]$ and $i \in [m_r - 1]$. We show that the next edge in $\mathcal{T}(S_2)$, i.e., $e_2 := (w_{\alpha-1}^{*2}, w_\alpha^{*2})$ is same as $e_1 := (w_{\alpha-1}^{*1}, w_\alpha^{*1})$, the next edge in $\mathcal{T}(S_1)$. The next edge can lead to one of the following cases:

1. Suppose e_1 leads to a dependent collision δ in S_1. Therefore the corresponding linear restriction E_δ must be spanned by $\mathcal{C}^{\mathrm{Acc}}(S_1^{\mathcal{E}_{\alpha-1}})$. Now $\mathcal{T}_{\alpha-1}(S_1) = \mathcal{T}_{\alpha-1}(S_2) \implies \mathcal{C}^{\mathrm{Acc}}(S_1^{\mathcal{E}_{\alpha-1}}) = \mathcal{C}^{\mathrm{Acc}}(S_2^{\mathcal{E}_{\alpha-1}})$. So E_δ is also spanned by $\mathcal{C}^{\mathrm{Acc}}(S_2^{\mathcal{E}_{\alpha-1}})$. As the message label is same, we must have $e_2 = e_1$.
2. Suppose e_1 leads to a new accident δ in S_1. As $\mathcal{C}^{Acc}(S_1) = \mathcal{C}^{Acc}(S_2)$, $\delta \in \mathcal{C}^{Acc}(S_2)$. Thus e_2 also leads to same accident δ in S_2. Thus $e_1 = e_2$.
3. Suppose e_1 leads to a a new vertex, i.e., $w_\alpha^{*1} \notin \mathcal{T}_{\alpha-1}(S_1)$. As the labels of both e_1 and e_2 are same, e_2 must also lead to a new vertex. Then using the definition of INS graph we have $e_1 = e_2$.

In all three cases we have $e_1 = e_2$, i.e., $\mathcal{T}_\alpha(S_1) = \mathcal{T}_\alpha(S_2)$. This proves the claim. So $|\mathsf{INstruct}^{\mathcal{M}})|$ is at most equal to the number of distinct accident basis of size a. Note that we can have at most $m = \sum_r m_r$ number of vertices in the graph. So the number of distinct accident basis is at most $\binom{m}{2}^a$. The result follows. ■

Corollary (Corollary 1)**.** *Let $a \geq 1$ be an integer. Then,*

$$\Pr[\mathbf{Acc}(\mathsf{INstruct}^{\mathcal{F}}) \geq a : \mathcal{F} \xleftarrow{\$} \mathsf{Func}] \leq \frac{m^{2a}}{N^a}.$$

Proof. The result can be derived by combining Lemma 3 and 4 followed by some algebraic simplifications. ■