A Unified Method for Improving PRF Bounds for a Class of Blockcipher based MACs

Mridul Nandi

National Institute of Standards and Technology
mridul.nandi@gmail.com

Abstract. This paper provides a unified framework for *improving* PRF (pseudorandom function) advantages of several popular MACs (message authentication codes) based on a blockcipher modeled as RP (random permutation). In many known MACs, the inputs of the underlying blockcipher are defined to be some deterministic affine functions of previously computed outputs of the blockcipher. Keeping the similarity in mind, we introduce a class of ADEs (affine domain extensions) and a wide subclass of SADEs (secure ADE) containing $\mathcal{C} = \{\mathsf{CBC\text{-}MAC}, \mathsf{GCBC}^*, \mathsf{OMAC}, \mathsf{PMAC}\}$. We define a parameter $N(t,q)$ for each domain extension and show that all SADEs have PRF advantages $O(tq/2^n + N(t,q)/2^n)$ where t is the total number of blockcipher computations needed for all q queries. We prove that PRF advantage of any SADE is $O(t^2/2^n)$ by showing that $N(t,q)$ is always at most $\binom{t}{2}$. We provide a better estimate $O(tq)$ of $N(t,q)$ for all members of $\mathcal{C}$ and hence these MACs have *improved advantages* $O(tq/2^n)$. Our proposed bounds for CBC-MAC and GCBC^* are better than previous best known bounds.

Keywords: affine domain extension, PRF, random permutation, CBC-MAC.

1 Introduction

Domain extension is a method to construct an extended function over an arbitrary domain when underlying function(s) over small domain are given. A common practice is to design domain extensions whose extended functions achieve some desired security whenever their underlying functions are assumed to have similar security. For example, it is well known that Merkle-Damgård with length strengthening padding [7, 16] extends a collision resistant compression function to a collision resistant hash function. Similarly MACs (message authentication codes) are also domain extensions extending small domain PRPs (pseudorandom permutations [13]) or PRFs (pseudorandom functions [8]) to arbitrary domain PRFs. A PRF and PRP have negligible advantage to be distinguished from the RF (random function) and RP (random permutation) respectively by any (q,t,ℓ)-*distinguisher* (*which makes q queries with ℓ and t invocations of RP to compute the output of the longest query and all queries respectively*). Any tuple of q such queries or messages are also called (q,t)- or (q,t,ℓ)-messages. In this paper we study MAC domain extensions based on a single blockcipher, modeled to be a RP on $\{0,1\}^n$ (or the Galois field $\mathbb{F}_{2^n}$ treated equally in the paper).

1.1 Related Works: PRF security analysis of Known MACs

The very basic and old domain extension method based on blockcipher is CBC[3] which was proven secure for prefix-free message spaces. Afterwards, many different variants of CBC are proven secure for arbitrary domains. In this paper we are mainly interested in the following domain extensions: $\mathcal{C} = \{$CBC-MAC [5], OMAC [9], GCBC* [18][1], PMAC [6]$\}$ and (directed acyclic graph) DAG-based PRFs [11, 19]. Our paper continues the following two lines of research which have been studied recently.

(1) Unifying Known Domain Extensions: In [11] a class of DAG based domain extensions was proposed where each *non-singular* DAG or a family of non-singular DAGs (see definition 2) corresponds to a domain extension. Even though the Jutla's class contains CBC, GCBC* and many other efficient domain extensions, it does not include those which encrypt a constant block (e.g. OMAC and PMAC encrypt the zero block). Each node of a DAG represents blockcipher invocation with input as a message block xor-ed with previously computed blockcipher-outputs corresponding to the predecessor nodes. If we add a special node representing to the encryption of the zero block then OMAC and PMAC can be included (as described in Nandi's class [19]).

(2)Finding Improved Bounds of PRF Advantages: The original PRF bound for the members of $\mathcal{C}$ and DAG-based constructions is $O(t^2/2^n)$ [3–6, 9–11, 19, 21]. The improved bound $O(\ell q^2/2^n)$ for CBC-MAC was shown in [2]. Afterwards, similar or better improved bounds were shown for other members of $\mathcal{C}$ [14, 15, 17] (except GCBC*) and for others (e.g. EMAC [21, 22], XCBC and TMAC [5, 12, 14]). See Table 1 for different known PRF bounds.

1.2 Motivation and Our Results

(1) We Unify Many Known Domain Extensions. In this paper we consider a more general class, called ADEs (*affine domain extensions*). It consists of all known domain extensions which invoke the underlying blockcipher π in a sequence such that inputs of π are determined from previous outputs of π via some affine functions. Moreover the output of the domain extension is the last output of the blockcipher. All members of it may not be secure (like we need non-singular DAG for PRF security). In section 3.2, we identify a class of SADEs (*secure* affine domain extensions) which are PRF secure and contains all modified non-singular DAG-based PRFs and all members of $\mathcal{C}$ (see theorem 2). The non-secure ADE does not necessarily mean insecure construction. We do not know any generic method to distinguish non-secure ADE. The other mentioned constructions such as EMAC, XCBC, TMAC, etc. do not directly fit into this class due to presence of one or more extra independent key (either auxiliary key or

[1] GCBC* is one example of one-key GCBC [18], a general class of CBC-type constructions which can include any number of keys. For simplicity, we only consider a particular one-key GCBC* which is eventually included in [11].

blockcipher key). They mostly have underlying CBC type structure for a single key. Generalized CBC class or GCBC includes those and considered in [18].

(2)We Find a PRF Bound for the Unified Class. Security analysis of all DAG-based constructions are based on the model that the underlying blockcipher is a random function and hence we can not go beyond $t^2/2^n$ bound in the RP-model of blockcipher due to the switching lemma [3] (switching from RF to RP costs $O(t^2/2^n)$). So we need to find different method to obtain improved bounds. The proof idea of [2] for the CBC-MAC uses structure graph whose nodes represent the internal outputs of the blockcipher and two successive outputs are connected by an arc. However, for a general ADE, one can not work simply with a graph since an input of blockcipher can be determined from more than one previous outputs. We use equivalence relation (denoted by $\sim$ with a possibly superscript) to capture the collisions on inputs of the blockcipher and use matrix (denoted by $\mathbf{A}$ with a possibly superscript) to capture the affine relation between the inputs and outputs of the blockcipher.[2] Theorem 3 shows that under some restriction on ℓ the advantage of any secure affine domain extension is bounded above by $(tq + N(t,q))/2^n$ where $N(t,q)$ is the maximum number of collision relations among all (q,t)-messages such that no pair of messages has accident two. We provide an informal definition of collision relation and number of accident. More detail definition can be found later.

1. A collision relation for a tuple of (q,t)-messages is a class of permutations which have identical collision patterns on the set of internal inputs of the blockcipher when we compute the domain extensions for the q messages. Mathematically, it is characterized by an equivalence relation on the set of indices of all intermediate inputs such that each collision pair is related.
2. Number of accident is the number of independent intermediate collisions which can not be derived from messages and other collisions only. All non-accident collisions are derived form the accidents (see definition 7 and remark 1 for more detail).

An easy upper bound t^2 of $N(t,q)$ can be shown for all SADEs and hence we have the generic classical PRF bound $O(t^2/2^n)$ for all SADEs (see theorem 4).

(3)We Find Improved PRF Bounds for CBC and GCBC*. Due to theorem 3, we only need to have a better estimation of $N(t,q)$ for a given secure affine domain extension. In section 7 we show that $N(t,q)$ is $O(tq)$ for each member of $\mathcal{C}$ and hence we obtain the improved PRF bounds $O(tq/2^n)$ for all members of $\mathcal{C}$ (see theorem 5). We do not know whether this upper bound holds for all secure affine domain extensions or not (this would be a challenging future work). Our improved bounds (see table 1 for comparison) are better than some of the previously known best bounds, namely $\ell q^2/2^n$ for CBC-MAC[2], and $t^2/2^n$ for GCBC*[18]. Note that the bound $\ell q^2/2^n$ can be worse compare to

[2] Clearly, inputs and outputs are related via blockcipher. Besides this relation, inputs of the blockcipher is affinely related (represented by a matrix) by the outputs. More importantly, this relation is independent of the blockcipher.

$tq/2^n$ or even $t^2/2^n$ if the query sizes are scattered enough. For example, when $\ell = q = t/2 = 2^{n/3}$ (this can happen if one message has ℓ blocks and all other messages have only one or two blocks) then $\ell q^2/2^n = 1$ and hence no security is guaranteed with $\ell q^2/2^n$ bound. On the other hand, $2qt/2^n = 4t^2/2^n = 2^{n/3}$ are negligible. So proving $t^2/2^n$ or $tq/2^n$ bound still guarantee the security.

Name of PRF	**Our PRF bounds**	Best Known Bounds	Other Bounds
CBC-MAC [3]	$\frac{\mathbf{11tq}}{\mathbf{2^n}}$ (R1)	$\frac{20\ell q^2}{2^n}$ (R1) [2]	$\frac{t^2}{2^n}$ [4, 19]
PMAC [6]	$\frac{\mathbf{5tq}}{\mathbf{2^n}}$ (R1)	$\frac{5tq}{2^n}$ [15]	$\frac{10\ell q^2}{2^n}$ [14]
OMAC [9]	$\frac{\mathbf{9tq}}{\mathbf{2^n}}$ (R1)	$\frac{5tq}{2^n}$ (R1) [17]	$\frac{3.5t^2}{2^n}$ [10]
GCBC* [18]	$\frac{\mathbf{11tq}}{\mathbf{2^n}}$ (R1)	$\frac{4t^2}{2^n}$ [18]	-
DAG-based [11, 19]	$\frac{\mathbf{t^2}}{\mathbf{2^{n-2}}}$	$\frac{t^2}{2^n}$ [11, 19]	-
SADE [this paper]	$\frac{\mathbf{3tq}}{\mathbf{2^n}} + \frac{\mathbf{N(t,q)}}{\mathbf{2^n}}$ (R1)	-	-

Table 1. PRF bounds for (q, t, ℓ)-distinguishers. $R1 : \ell < 2^{n/3-1}$, $R2 : \ell < \min\{2^{n/8}, q^{1/2}\}$ and see section 6 for $N(t, q)$.

2 Preliminaries

2.1 Notation and Convention

$\mathbb{F}_{2^n}$ is the finite field $\{0,1\}^n$ with "+" and "$\cdot$". Here as a set $\mathbb{F}_{2^n}$ and $\{0,1\}^n$ are same and elements of it are called blocks. The blocks $\mathbf{0}$, $\mathbf{1}$ are additive and multiplicative identities respectively. We denote a permutation by π and the set of all permutations on $\{0,1\}^n$ by $\mathbb{P}_n$. The notation $\Pi \xleftarrow{*} \mathbb{P}_n$ means that Π is chosen randomly from $\mathbb{P}_n$ and is called random permutation. The number of elements of S is denoted by $\#S = s$ (say) and we call S to be an s-set. Similarly we also call t-tuple or t-vector (e.g. $\mathbf{y} = (y(1), \ldots, y(t))$ and $(s \times (s+1))$-matrix etc. We write $\mathbf{P}(m, r) = m(m-1)\ldots(m-r+1)$. We denote $(i,j)^{\text{th}}$ entry and i^{th} row of a $s \times (s+1)$-matrix $\mathbf{A}$ by $a_{i,j}$ and A_i respectively, $1 \le i \le s, 0 \le j \le s$.

The Domain and range of a function $g : J \to \mathbb{F}_{2^n}$ are J and $g(J) := \{g(j) : j \in J\}$ respectively and denoted by $D(g)$ and $R(g)$ respectively. If $J' \subseteq J$ then $g(J')$ is the range of $g|_{J'}$ restricted on the domain J'. If $f : J \to \mathbb{F}_{2^n}$ then $\pi(f) : J \to \mathbb{F}_{2^n}$ is the function such that $\pi(f)(j) = \pi(f(j))$ for all $j \in J$. If $f(i) = a_{i,0} + \sum_{j=1}^{i-1} a_{i,j} \cdot f(j), 1 \le i \le s$ for some $g : [1, s] \to \mathbb{F}_{2^n}$ then f is called the corresponding function of g and it is denoted by $g \rightarrowtail_{\mathbf{A}} f$. We denote the functions having domain $[1, t] := \{1, 2, \ldots, t\}$ (index-set) by x, y, f, g etc. We

denote an equivalence relation[3] on $[1,t]$ by $\sim$. The relation $\sim_i$ is $\sim$ concentrated on $[1,i]$ i.e. $j \sim_i j'$ if and only if either $j = j'$ or $j \sim j'$ and $j, j' \leq i$. The equivalence class containing i is $[[i]]$ and the minimum element of the class is called leader. The set of all leaders is denoted by $Ld(\sim) := \{\imath_1, \ldots, \imath_s\}$. For any function $g : J \to \mathbb{F}_{2^n}$, the induced equivalence relation $\sim^g$ is defined as $i \sim^g j$ if and only if $g(i) = g(j)$. Two function f and g of same domain are said to be equality-matching if $\sim^f = \sim^g$ and we denote it by $f \stackrel{\circ}{=} g$.

The messages from the message space $\mathcal{M}$ are denoted by M, M_i, etc. $\ell := \ell(M)$ denote the number of π-invocations to compute the output of the domain extension $\mathcal{D}^\pi(M)$. The query tuple $(M_1, \ldots, M_q)$ and the response tuple $(w_1, \ldots, w_q)$ are denoted by $\mathbf{M}$ and $\mathbf{w}$ respectively where q is the number of queries. The final and intermediate index-sets are $I = \{t_1, t_2, \ldots, t_q\}$ and $\overline{I} := [1,t] \setminus I$ respectively where $\ell_i = \ell(M_i)$, $t_i := \sum_{j=1}^{i} \ell_j$, $t := t_q$.

2.2 Decorrelation Theorem

We first state a useful result (lemma 22 of [25]) for PRF security analysis[4] and we call it *Decorrelation Theorem*. The main idea of the theorem was described as Patarin's "coefficient H-techniques" [20] (according to Vaudenay [25, 24]). Different generalized versions are stated in [4, 19]. For a deterministic adaptive distinguisher, the queries and even the number of blocks of queries may be dependent random variables. The decorrelation theorem gets rid of the correlation and reduces PRF security analysis of $\mathcal{D}$ to show that the q-decorrelation probability $\mu_{\mathbf{M},\mathbf{w}} := \mathsf{Pr}[\mathcal{D}^{\Pi}(M_1) = w_1, \ldots, \mathcal{D}^{\Pi}(M_q) = w_q : \Pi \stackrel{*}{\leftarrow} \mathbb{P}_n]$ is very close to $\frac{1}{2^{nq}}$ (the corresponding q-decorrelation probability for RF) where $\mathbf{M} = (M_1, \ldots, M_q) \in \mathcal{M}^q$ and $\mathbf{w} = (w_1, \ldots, w_q) \in \mathbb{F}_{2^n}^q$ two q-tuples of distinct elements. We call such $\mathbf{M}$ and $\mathbf{w}$ coordinate-wise distinct. We denote the set $\{w_1, \ldots, w_q\}$ by W.

A big advantage in the probability computation is that the source of randomness is only from the uniform distribution of Π over $\mathbb{P}_n$. The intuitive reason why it works for bounding PRF advantage is the following: Any adaptive distinguisher eventually makes decision based on all queries and responses. So if for any possible set of queries, the responses of $\mathcal{D}$ is almost uniformly random then no adaptive distinguisher can distinguish it from a random function with non-negligible probability. Let RF be the random function from $\mathcal{M}$ to $\mathbb{F}_{2^n}$. The distinguishing advantage of a domain extension $\mathcal{D}$ over a message space $\mathcal{M}$ based on a random permutation Π is defined as follows:

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{\Pi}}(\mathcal{A}) = \mathsf{Pr}[\mathcal{A}^{\mathsf{RF}} = 1] - \mathsf{Pr}[\mathcal{A}^{\mathcal{D}^{\Pi}} = 1], \text{ and } \mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{\Pi}}(q,t,\ell) = \max_{\mathcal{A}} \mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{\Pi}}(\mathcal{A})$$

where maximum is taken over all (q,t,ℓ)-distinguishers.

[3] A Binary relation is a subset of $[1,t]^2$ and a member (i,j) is denoted as $i \sim j$. An equivalence relation is a binary relation on $[1,t]$ satisfying the reflexive $(i \sim i, \forall i)$, symmetric $(i \sim j \Rightarrow j \sim i)$ and transitive $(i \sim j, j \sim k \Rightarrow i \sim k)$.

[4] The technique is also applicable for (strong) pseudorandom permutation [23], pseudo online cipher [19], etc.

Theorem 1. (Decorrelation Theorem)
Let q, t and ℓ be fixed integers, ϵ be some positive real number (may depend on q, t, ℓ) and $\mathcal{D}^{\pi} : \mathcal{M} \to \mathbb{F}_{2^n}$ be a domain extension, $\pi \in \mathbb{P}_n$ such that $\mu_{\mathbf{M},\mathbf{w}} \geq (1-\epsilon) \times 2^{-nq}$ for all coordinate-wise distinct $\mathbf{M}, \mathbf{w}$ with $\sum_{i=1}^{q} \ell(M) \leq t$ and $\max_i \ell(M_i) \leq \ell$. Then $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{\Pi}}(\mathcal{A}) \leq \mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{\Pi}}(q, t, \ell) \leq \epsilon + \frac{q(q-1)}{2^{n+1}}$ $\forall (q, t, \ell)$-distinguisher $\mathcal{A}$.

The proof of the above theorem is given in Appendix. Security analysis of PRF base on any blockcipher E_K is same if we incorporate the PRP advantage of the E_K by using the well known hybrid argument technique. By using hybrid technique it is well known that $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{E_K}}(q, t, \ell) \leq \mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{\Pi}}(q, t, \ell) + \mathbf{Adv}^{\mathrm{prp}}_{E_K}(t)$.

3 Affine Domain Extension

A keyed blockcipher is nothing but a permutation π. To reduce key-size it may be desired to design a single key based domain extension. Clearly a reasonable domain extension based on a single permutation π must invoke the permutation π several times. To make it efficient, the inputs (called intermediate inputs) to π should be computed via some simple functions. The affine domain extensions (or ADEs) are permutation-based domain extensions where the intermediate inputs are determined by some affine functions of outputs of π (called intermediate outputs) and the final outputs of the domain extensions are the last outputs of π. Message play role in defining the intermediate inputs. We first study CBC which is a simple example of ADE.

Example 1. The cipher block chaining or CBC is defined as $\mathsf{CBC}^{\pi}(M) = \pi(\ldots \pi(\pi(\alpha_1)+\alpha_2)\ldots+\alpha_b)$ where $M = (\alpha_1, \ldots, \alpha_b) \in \mathbb{F}_{2^n}^b$. If the i^{th} intermediate input and output of π are $x(i)$ and $y(i)$ respectively then we have

$$x(1) = \alpha_1,\ y(1) = \pi(x(1)), x(i) = y(i-1) + \alpha_i,\ y(i) = \pi(x(i))\ 2 \leq i \leq b. \quad (1)$$

For a fixed message M, $x(i)$'s are affine functions of $y(i)$'s. We call the functions x and y input and output functions respectively for the message M and the permutation π. The i^{th} intermediate input and output are $x(i)$ and $y(i)$ respectively. We represent the above relation by a lower triangular $(b \times (b+1))$-matrix[5] $\mathbf{A}_{\mathsf{CBC}} = ((a_{i,j}))$ whose rows are the followings:

-the first row: $\mathbf{A}_1 = (\alpha_1, \mathbf{0}, \ldots, \mathbf{0})$

-the i^{th} row: $\mathbf{A}_i = (\alpha_i, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{1}, \mathbf{0}, \ldots, \mathbf{0})$ for $i \geq 2$, where $\mathbf{1}$ appears on $(i-1)^{\text{st}}$ column (starting with zeroth column).

The $(i,j)^{\text{th}}$ element $a_{i,j}$ corresponds to the coefficient which is multiplied with $y(j)$ to define $x(i)$, $1 \leq i, j \leq b$. The $(i,0)^{\text{th}}$ element $a_{i,0}$ corresponds to the i^{th} message block (starting from 1^{st} message block). Note that the matrix $\mathbf{A}$ depends on the message only. Thus, CBC can be characterized by a family of matrices (which we call coefficient matrices) indexed over a message space. For any message M, CBC can be defined as $\mathsf{CBC}(M) = y(b)$ where $\mathbf{A} = \mathbf{A}_{\mathsf{CBC}} := ((a_{i,j}))$ is the coefficient matrix for the message M and

[5] A matrix $\mathbf{A}_{s \times (s+1)} = ((a_{i,j}))$ is called *lower triangular* if $a_{i,j} = \mathbf{0}$, $1 \leq i < j \leq s$.

E1- $y \rightarrowtail_{\mathbf{A}} x:\ x(i) = a_{i,0} + \sum_{j=1}^{i-1} a_{i,j} \cdot y(j),\ 1 \le i \le b$,
E2- $y = \pi(x):\ y(i) = \pi(x(i)), 1 \le i \le b$.

Note that for any matrix $\mathbf{A}_{b\times(b+1)}$ (not necessarily the coefficient matrix of CBC) and a permutation π the values $x(1), y(1), x(2), y(2), \ldots$ are uniquely determined which satisfy E1 and E2 as described above. More precisely, $x(i)$ is uniquely determined from $y(1), \ldots, y(i-1)$ and $\mathbf{A}$ and $y(i)$ is uniquely determined from π and $x(i)$. We may denote the pair of the unique solution by $(x^{\pi,\mathbf{A}}, y^{\pi,\mathbf{A}})$ or (x^{π}, y^{π}) (when $\mathbf{A}$ is understood). These are called *input and output functions induced by* π.

Definition 1. (Affine Domain Extension or ADE).
A domain extension $\mathcal{D}$ is called ADE over a message space $\mathcal{M}$ if for each message $M \in \mathcal{M}$ there is a lower triangular matrix $\mathbf{A}_{\ell\times(\ell+1)}$ (called coefficient matrix corresponding to M) for some $\ell = \ell(M)$ such that $\mathcal{D}^{\pi}(M) = y^{\pi,\mathbf{A}}(\ell)$, for all $\pi \in \mathbb{P}_n$.

The elements $a_{i,j}$'s of $\mathbf{A}$ may be message blocks or constants, such as $\mathbf{0}$ and $\mathbf{1}$, or some constants depending on message type such as padding is applied or not. These elements are independent of the permutation π. This representation is useful when we study a domain extension for a fixed message with different choices of permutations.

Collision relation: For any fixed $\mathbf{A}$, the equivalence relation induced (see section 2.1) by x^{π} and y^{π} are same and by abuse of notation we denote the common relation by $\sim^{\pi} := \sim^{y^{\pi}}$. We call it a *collision relation* as it captures all collisions on outputs of x and y. In [2] a structure graph is used to capture all collisions.

3.1 Examples of Affine Domain Extensions

The cipher block chaining message authentication code or CBC-MAC [3, 21] is a very basic and old method to extend the domain of PRF. Later, many CBC-type domain extensions were proposed. For a message M, let $M^* = M\|10^d$ with the smallest nonnegative d so that $n \mid |M^*|$ (n divides $|M^*|$). If $n \mid |M|$ then $\delta = 0$, $\overline{M} = M$, otherwise $\delta = 1$ and $M = M^*$. We represent $\overline{M}$ and M^* by $(\alpha_1, \ldots, \alpha_b) \in \mathbb{F}_{2^n}^b$. The integer $b := b(M) = \lceil |M|/n \rceil$ is called the *number of blocks* of M. The keyed blockcipher is denoted by $\pi \in \mathbb{P}_n$. We show some CBC-type domain extensions such as CBC-MAC, GCBC*, OMAC, and others such as PMAC, DAG-based PRF are affine domain extensions. The definitions of these are based on some distinct non-$\mathbf{0}$, non-$\mathbf{1}$ constants c_i''s and c_δ such that their differences are not $\mathbf{1}$. The original choices of constants can be found in their respective papers [6, 9, 18]. In the following, we define $y(i) = \pi(x(i))$.

CBC-MAC [3]: In example 1 we have defined CBC for messages of size multiple of n. For a general message one can use some padding rule. One such example is CBC-MAC$^{\pi}(M)$ = CBC$^{\pi}(M^*)$. So CBC-MAC is nothing but CBC applied to the padded message. Let $\ell(M) = b$ and the input function $x(1) = \alpha_1$ and $x(i) = \alpha_i + y(i-1),\ \ 2 \le i \le b$.

$$\mathbf{A}_1 = \begin{pmatrix} \alpha_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_2 & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_3 & \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \alpha_b & \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} & \mathbf{0} \end{pmatrix} \quad \mathbf{A}_2 = \begin{pmatrix} \alpha_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_2 & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_3 & \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \alpha_b & \mathbf{0} & \mathbf{0} & \dots & c_\delta & \mathbf{0} \end{pmatrix} \quad \mathbf{A}_3 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_2 & \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \alpha_{b+1} & c_\delta & \mathbf{0} & \dots & \mathbf{1} & \mathbf{0} \end{pmatrix}$$

$$\mathbf{A}_4 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_1 & c'_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \alpha_{b-1} & c'_{b-1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \alpha_b & c_\delta & \mathbf{1} & \dots & \mathbf{1} & \mathbf{0} \end{pmatrix}$$

Fig. 1. $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4$ are the coefficient matrices of CBC, GCBC, OMAC and PMAC respectively for the message M.

GCBC [18]: In case of GCBC, we consider the messages with $b \geq 2$ and it is defined as $\mathsf{GCBC}^\pi(M) = \pi\big(\alpha_b + c_\delta \cdot \mathsf{CBC}^\pi(\alpha_1, \dots, \alpha_{b-1})\big)$ for some constants c_0 and c_1. The input function is same as CBC-MAC except the final intermediate input $x(b) = \alpha_b + c_\delta \cdot y(b-1)$.

OMAC [9]: $\mathsf{OMAC}^\pi(M) = \pi\big(\alpha_b + c'_\delta \cdot \pi(\mathbf{0}) + \mathsf{CBC}^\pi(\alpha_1, \dots, \alpha_{b-1})\big)$ where $\mathsf{CBC}^\pi(\lambda) = \mathbf{0}$, λ is the empty string. Let $\ell(M) = b+1$ and the input function is

$$x(1) = \mathbf{0},\ x(i) = \alpha_{i-1} + y(i-1),\ 2 \leq i < b+1 \text{ and } x(b+1) = \alpha_b + c_\delta \cdot y(1) + y(b).$$

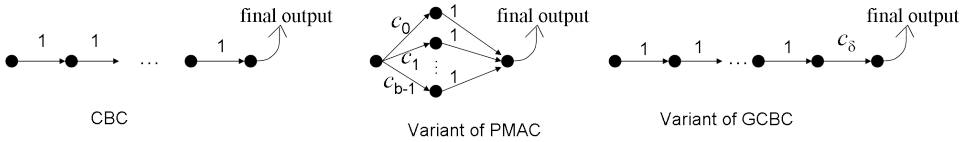

Fig. 2. The DAG representation of CBC, a variant of PMAC and GCBC.

PMAC [6]: $\mathsf{PMAC}^\pi(M) = \pi(\alpha_b + \sum_{i=1}^{b-1} \pi(\alpha_i + c'_i \cdot \pi(\mathbf{0})) + c_\delta \cdot \pi(\mathbf{0}))$. So $\ell(M) = b+1$ and the input function $x(1) = \mathbf{0},\ x(i) = \alpha_{i-1} + c'_i \cdot y(1), 2 \leq i < b$, and $x(b+1) = \alpha_b + c_\delta \cdot y(1) + \sum_{i=2}^{b} y(i)$.

DAG-based PRF [11, 19]: In [11, 19] a domain extension over a message space $\mathcal{M} = \mathbb{F}_{2^n}^b$ is proposed for every non-singular labeled DAG $G = ([1, b], \mathcal{E}, c)$ where $\mathcal{E}$ is the set of arcs and $c : \mathcal{E} \to \mathbb{F}_{2^n}$ corresponds to the label. In [11], a more general domain extension is defined for arbitrary messages by considering a family of DAGs where each DAG corresponds to the domain extension with fixed length messages after padding. The general definition includes CBC-MAC for arbitrary message space, the version of GCBC considered in our paper. In [19], a much bigger class is considered which can include PMAC and OMAC. All these constructions are affine domain extensions. Here we show it for the construction based on a labeled DAG with message space $\mathcal{M} = \mathbb{F}_{2^n}^b$ and leave readers to verify for other cases.

Definition 2. *A DAG G with b nodes $[1, b]$ is called non-singular [11] if the exists exactly one source node (in-degree is zero), one sink node b (out-degree is zero) and for any two nodes v and v' with same set of incident nodes U (i.e. $U = \{u : u \to v\} = \{u : u \to v'\}$), there exists $u \in U$ such that $c(u, v) \neq c(u, v')$.*

The nodes are numbered in such a way that $u \to v$ implies $u < v$. This is possible since G has no cycle. Given a message $M = (\alpha_1, \ldots, \alpha_b) \in \mathbb{F}_{2^n}^b$, let $\ell = \ell(M) = \ell$ and $\mathsf{DAG}_G(M) = y(\ell)$ where the input and output functions are

$$\mathsf{DAG}_G(M) = y(\ell), \text{ where } x(v) = \alpha_v + \sum_{v' \to v} c(v', v) \cdot y(v'),\ y(v) = \pi(x(v)), 1 \leq v \leq \ell.$$

Hence any DAG-based domain extension is ADE. The $(i, j)^{\text{th}}$ entry of the coefficient matrix is $a_{i,j} = c(i, j)$ if $i \to j$, otherwise $a_{i,j} = \mathbf{0}$. The $(i, 0)^{\text{th}}$ entry is the the i^{th} message block α_i. It is easy to verify the following result.

Lemma 1. *If a DAG G is non-singular then for any message all rows of the coefficient matrix are distinct. If $M \neq M'$ then $\mathbf{A}_\ell^M \neq \mathbf{A}_i^{M'}$, $1 \leq i \leq \ell$.*

EMAC [21], XCBC [5], TMAC [12] (as these domain extensions require either auxiliary keys or more than one permutation) and XOR-MAC [1] (the output is sum of all previous intermediate outputs instead of the last intermediate output) are some examples of non-ADE PRFs.

3.2 Secure Affine Domain Extension

In this section we characterize a class of PRF secure affine domain extension called secure affine domain extensions.

Definition 3. (Secure Affine Domain Extension or SADE). *An ADE $\mathcal{D}$ is called SADE if for any $(i, M) \neq (\ell(M'), M')$, $\exists \pi \in \mathbb{P}_n$ such that $y^{\pi, M}(i) \neq \mathcal{D}^\pi(M')$. In other words, an affine domain extension $\mathcal{D}$ is non-secure if $\exists (i, M) \neq (\ell(M'), M')$ such that $y^{\pi, M}(i) = \mathcal{D}^\pi(M')$, $\forall \pi \in \mathbb{P}_n$.*

Informally speaking, an ADE $\mathcal{D}$ is non-secure (not necessarily insecure) if $\mathcal{D}^\pi(M)$ always collide with a specific intermediate output of π while computing $\mathcal{D}^\pi(M')$ for some messages M and M'. We call this type of collision "forced collision" (later we see that it is related to a special collision relation called forced collision relation). By knowing the value of $\mathcal{D}^\pi(M')$ of a non-secure ADE (even for secretly chosen permutation π), a specific positioned intermediate output of $\mathcal{D}^\pi(M)$ is leaked. This may be an undesired property which could lead a distinguishing attack. For example, we have the following attacks:

1. CBC-MAC on $\{0, 1\}^*$ is not SADE and it has length extension attack due to this forced collision.
2. If we modify the definition of OMAC by choosing $c_0 = \mathbf{1}$ then $\mathcal{D}(\mathbf{0}, \mathbf{0}) = \pi(\mathbf{0}) = y^{\pi, M}(1)$ for all permutation π and a message M (if we set $x^{\pi, M}(1) = \mathbf{0}$). So it is not a SADE and one can show a distinguishing attack exploiting this observation (e.g., $\mathcal{D}(\mathbf{0}, \mathbf{0}) = C \Rightarrow \mathcal{D}(C) = C$ with probability one). It also explains why we should choose non-$\mathbf{1}$ constants for OMAC.

Even though we know some attacks on non-secure ADE, we do not know yet how to make a generic attack on all non-secure constructions. All affine domain extensions avoiding this undesired forced collisions is SADE. Now we provide an equivalent definition of secure affine domain extensions which can be used to verify the known domain extensions are secure. We first define a **joint coefficient matrix** which is nothing but a proper combination of coefficient matrices. Let $\mathcal{D}$ be a domain extension over a message space $\mathcal{M}$. We fix two q-tuples $\mathbf{M} = (M_1, \ldots, M_q) \in \mathcal{M}^q$ and $\mathbf{w} = (w_1, \ldots, w_q) \in \mathbb{F}_{2^n}^q$ such that M_i's and w_i's are distinct. We denote $I := I_{\text{final}} = \{t_1, t_2, \ldots, t_q := t\}$, $\overline{I} = [1, t] \setminus I$ where $t_i = \sum_{j=1}^{i} \ell_j$ and $\ell_j := \ell(M_j)$. Like coefficient matrix, the joint computations of $\mathcal{D}^\pi$ (i.e. $\mathcal{D}^\pi(M_1), \ldots \mathcal{D}^\pi(M_q)$) can also be represented by a single combined matrix, called joint coefficient matrix. It is a proper combination of coefficient matrices $\mathbf{A}^{M_i} := (\mathbf{b}_i : \mathbf{C}_i)$, where $\mathbf{b}_i$ is a ℓ_i-vector and $\mathbf{C}_i$ is a $\ell_i \times \ell_i$ matrix, $1 \le i \le q$. The joint coefficient matrix $\mathbf{A}^{\mathbf{M}}$ is defined below.

$$\mathbf{A}^{\mathbf{M}}_{t\times(t+1)} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{C}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{b}_2 & \mathbf{0} & \mathbf{C}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{b}_q & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{C}_q \end{pmatrix}$$

It is easy to see that the joint coefficient matrix $\mathbf{A}$ is a lower triangular matrix with $a_{i,j} = \mathbf{0}$, $\forall j \in I$ and $\forall i$ (i.e. j^{th} column is zero vector for all $j \in I$). Let (x^π, y^π) be the unique solution of E1 and E2 (i.e. $\pi(x^\pi) = y^\pi$ and $y^\pi \rightarrowtail x^\pi$). Then we can prove that $y^\pi(t_i) = \mathcal{D}^\pi(M_i)$ for all permutation π and $1 \le i \le q$.

Lemma 2. *Let (x, y) and (x_i, y_i) be the unique solution of E1 and E2 for the permutation π and the joint coefficient matrix $\mathbf{A}^{\mathbf{M}}$ (defined above) and A^{M_i} respectively. Then, $\mathcal{D}^\pi(M_i) = y_i(\ell_i) = y(t_i)$ More generally*

$$x(t_{i-1} + j) = x_i(j),\ y(t_{i-1} + j) = y_i(j),\ 1 \le j \le q,\ 1 \le i \le q. \tag{2}$$

The result follows trivially by showing that the pair of functions (x, y) defined in eq. 2 is the unique solution of E1 and E2 for the matrix $\mathbf{A}^{\mathbf{M}}$.

Lemma 3. Equivalent definition of SADE
A affine domain extension $\mathcal{D}$ is secure if and only if for any messages $M_1 \neq M_2$ and $i \neq \ell_1 + \ell_2 := t$, there exists a permutation π such that $x^{\pi,\mathbf{A}}(i) \neq x^{\pi,\mathbf{A}}(t)$ for the induced input function $x^{\pi,\mathbf{A}}$ where $\mathbf{A}$ is the joint coefficient matrix for the tuple (M_1, M_2).

4 Some Useful Results on Output functions and Collision Relations

In this section, we fix a lower triangular matrix $\mathbf{A}_{t\times(t+1)} = ((a_{i,j}))$ for some integer $t > 0$, usually it is either a coefficient matrix of a message M (see definition 1) with $t = \ell = \ell(M)$ or a joint coefficient matrix for a tuple of q distinct messages (see in section 3.2) $(M_1, \ldots, M_q)$ with $t = \sum_{i=1}^{q} \ell(M_i)$. We refer readers to section 2.1 for meaning of notation whenever needed.

4.1 Output Function

While computing $\mathcal{D}^\pi(M_1), \ldots, \mathcal{D}^\pi(M_q)$, the permutation π is invoked successively. Suppose $x(1), \ldots, x(t)$ are all inputs to π and their corresponding outputs are $y(i) = \pi(x(i))$, $1 \le i \le t$. The sequence of the inputs and outputs can be characterized by input and output functions x and y respectively. In other words, outputs of the functions x and y are nothing but the inputs and outputs of π while computing $\mathcal{D}^\pi(M_1), \ldots, \mathcal{D}^\pi(M_q)$.

Definition 4. *Let $\mathbf{A}^{\mathbf{M}}_{t\times(t+1)}$ be the joint coefficient matrix for $(M_1, \ldots, M_q)$. To $\mathbf{M}$ and any $\pi \in \mathbb{P}_n$ we associate an input and output functions $x, y : [1, t] \to \mathbb{F}_{2^n}$ where $y \rightarrowtail_{\mathbf{A}} x$ (i.e. $x(i) = \sum_{j=0}^{i-1} a_{i,j} \cdot y(j), \forall i$) and $\pi(x) = y$ (i.e. $\pi(x(i)) = y(i), \forall i$). We denote the output function by $y^{\pi,\mathbf{A}}$ or y^π (as $\mathbf{A}$ is fixed).*

For any function $g : [1, t] \to \mathbb{F}_{2^n}$, let $\mathbb{P}_n[g]$ denote the set of all permutations with g as an output function. That is, $\mathbb{P}_n[g] = \{\pi : y^\pi = g\}$. The following results describe an equivalent characterization of a output function and compute the number of permutations which induce the output function. The proofs are straightforward and hence we skip the proof. Recall that $x \stackrel{\circ}{=} y$ if $x(i) = x(j)$ if and only if $y(i) = y(j)$ (i.e. the collision patterns of x and y are same).

Lemma 4. (characterization of an output function) *Given any function $y : [1, t] \to \mathbb{F}_{2^n}$ with $s = \#R(y)$ (the size of range-set of y). Then,*

$$\#\mathbb{P}_n[y] = \begin{cases} (2^n - s)! & \text{if } x \stackrel{\circ}{=} y,\ y \rightarrowtail x \\ 0, & \text{otherwise} \end{cases}$$

So y is an output function if and only if $y \stackrel{\circ}{=} x$ where $y \rightarrowtail x$.

4.2 Collision Relation

In the previous subsection we have defined input and output function. A collision relation is the equivalence relation capturing the collision pattern of them.

Definition 5. *An equivalence relation $\sim$ is called collision relation (w.r.t. $\mathbf{A}$) if there exists a permutation π such that $i \sim j$ if and only if $y(i) = y(j)$ (i.e. $\sim^\pi := \sim^{y^{\pi,\mathbf{A}}} = \sim$). Let $\mathbb{P}_n[\sim] := \{\pi : \sim^\pi = \sim\}$.*

Any equivalence relation on $[1, t]$ is not necessarily a collision relation. In this section we provide a characterization of collision relations. For any $(t+1)$-vector $\mathbf{v} = (v_0, v_1, \ldots, v_t) \in \mathbb{F}_{2^n}^{t+1}$ and any arbitrary equivalence relation $\sim$ we define a $\sim$-reduced vector $\mathbf{v}^\sim = (v_0, v_1^\sim, \ldots, v_t^\sim)$ and the following sets of $(t+1)$-vectors.

1. $\mathcal{V}_{eq} := \{\mathbf{A}_i^\sim - \mathbf{A}_j^\sim : i \sim j\}$,
2. $\mathcal{V}_{neq} := \{\mathbf{A}_i^\sim - \mathbf{A}_j^\sim : i \not\sim j\}$
3. $\mathcal{V}^*_{neq} := \mathcal{V}_{neq} \cup \{\mathbf{e}_i - \mathbf{e}_j : i \not\sim j\}$
4. $\mathcal{V}^{**}_{neq} := \mathcal{V}^*_{neq} \cup \{\mathbf{e}_0\}$

where $\mathbf{A}_i$ is the i^{th} row of $\mathbf{A}$, $\mathbf{e}_k \in \mathbb{F}_{2^n}^{t+1}$ is the $(t+1)$-vector whose k^{th} entry is $\mathbf{1}$ and all others are $\mathbf{0}$, $0 \le k \le t$, and

$$v_i^\sim = \begin{cases} \sum_{j \in [[i]]} v_j, & \text{if } i \in Ld(\sim) \\ \mathbf{0}, & \text{o.w.} \end{cases}$$

Let $\overline{\mathbf{y}} = (\mathbf{1}, y(1), \ldots, y(t))$ be the vector for an output function y with collision relation $\sim$ then $(\mathbf{A}_i^{\sim} - \mathbf{A}_j^{\sim}) \cdot \overline{\mathbf{y}} = x(i) - x(j) = \mathbf{0}$ whenever $i \sim j$. Thus, $\forall \mathbf{v} \in \mathcal{V}_{eq}$, $\mathbf{v} \cdot \overline{\mathbf{y}} = \mathbf{0}$. Similarly, one can prove that $\forall \mathbf{v} \in \mathcal{V}_{neq}^{**}$, $\mathbf{v} \cdot \overline{\mathbf{y}} \neq \mathbf{0}$.

Example 2. This is an example considered for CBC in [2]. Now we revisit it in our joint coefficient matrix notations. Let $M = (\alpha_1, \alpha_2, \alpha_3)$ and $M' = (\alpha_1', \alpha_2', \alpha_3')$ such that $\alpha_1 \oplus \alpha_3 = \alpha_1' \oplus \alpha_3'$. Now, consider a relation $\sim = \{\{1,6\},\{2,5\},\{3,4\}\}$. Thus, $Ld(\sim) = \{1,2,3\}$. The coefficient matrix $\mathbf{A} = \mathbf{A}^{M,M'}$ of CBC and the reduced matrix $\mathbf{A}^{\sim}$ are computed below:

$$\mathbf{A}^{M,M'} = \begin{pmatrix} \alpha_1 & \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_2 & \mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_3 & \mathbf{0}\,\mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_1' & \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_2' & \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{1}\,\mathbf{0}\,\mathbf{0} \\ \alpha_3' & \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{1}\,\mathbf{0} \end{pmatrix}, \quad \mathbf{A}^{\sim} = \begin{pmatrix} \alpha_1 & \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_2 & \mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_3 & \mathbf{0}\,\mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_1' & \mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_2' & \mathbf{0}\,\mathbf{0}\,\mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \\ \alpha_3' & \mathbf{0}\,\mathbf{1}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0}\,\mathbf{0} \end{pmatrix}$$

Note that $\mathbf{A}_1^{\sim} + \mathbf{A}_6^{\sim} = \mathbf{A}_3^{\sim} + \mathbf{A}_4^{\sim}$ and hence the collision $y_1 = y_6$ is determined by the collision $y_3 = y_4$. The set $\mathcal{V}_{eq} = \{\mathbf{A}_i^{\sim} - \mathbf{A}_j^{\sim} : i \sim j\}$ has only two independent vectors. So the rank for the relation is two, even though it has three pairs which are related (it is termed as true collision in [2]).

Lemma 5. Characterization of Collision Relation
Let $\sim$ be a collision relation then there exists $y : [1,t] \to \mathbb{F}_{2^n}$ such that
N1: $v_0 + \sum_{j=1}^{t} v_j \cdot y(j) = \mathbf{0}$ for all $(v_0, v_1, \ldots, v_t) \in \mathcal{V}_{eq}$,
N2: $v_0 + \sum_{j=1}^{t} v_j \cdot y(j) \neq \mathbf{0}$ for all $(v_0, v_1, \ldots, v_t) \in \mathcal{V}_{neq}^{**}$.
Hence, a necessary condition for a collision relation is that each vector of $\mathcal{V}_{neq}^{**}$ is linearly independent with $\mathcal{V}_{eq}$. Conversely, if $\mathcal{V}_{neq}^{**}$ is linearly independent with $\mathcal{V}_{eq}$ (i.e. $\sim$ satisfies the above necessary condition) then

$$(2^n - s)! \times 2^{n(s-a)} \times (1 - \frac{\#\mathcal{V}_{neq}^{*}}{2^n}) \;\leq\; \#\mathbb{P}_n[\sim] \;\leq\; (2^n - s)! \times \mathbf{P}(2^n, s-a)$$

where $s = \#Ld(\sim)$ and $a = \mathsf{acc}(\sim) := \mathrm{rank}(\mathcal{V}_{eq})$. Hence $\sim$ is a collision relation if $\#\mathcal{V}_{neq}^{*} < 2^n$.

Proof. If we know that $\sim$ is an equivalence relation induced by the vector $(y(1), \ldots, y(t))$ then for any vector $v = (v_0, \ldots, v_t)$ we have $v_0 + \sum_{i=1}^{t} v_i \cdot y(i) = v_0 + \sum_{i=1}^{t} v_i^{\sim} \cdot y(i)$. If $y \to_{\mathbf{A}} x$ then $x(i) = \mathbf{A}_i \cdot (\mathbf{1}, y(1), \ldots, y(t)) = \mathbf{A}_i^{\sim} \cdot (\mathbf{1}, y(1), \ldots, y(t))$. If y is an output function then $\sim^x = \sim$ and hence $i \sim j$ if and only if $x(i) = x(j)$. In other words, $v_0 + \sum_{j=1}^{t} v_j \cdot y(j) = \mathbf{0}$ for all $(v_1, v_1, \ldots, v_t) \in \mathcal{V}_{eq}$, and $v_0 + \sum_{j=1}^{t} v_j \cdot y(j) \neq \mathbf{0}$ for all $(v_0, v_1, \ldots, v_t) \in \mathcal{V}_{neq}$. Moreover, $y(i) \neq y(j)$ whenever $i \not\sim j$. So $v_0 + \sum_{j=1}^{t} v_j \cdot y(j) \neq \mathbf{0}$ for all $(v_0, v_1, \ldots, v_t) \in \mathcal{V}_{neq}^{**}$. So each vector of $\mathcal{V}_{neq}^{**}$ is linearly independent with $\mathcal{V}_{eq}$. If not, let $v \in \mathcal{V}_{neq}^{**}$ such that $v = \sum_{i=1}^{r} c_i \cdot u^i$ for some constants $c_i \in \mathbb{F}_{2^n}$ where $u^i \in \mathcal{V}_{eq}$. Since the vector product $u^i \cdot (1, y(1), \ldots, y(t)) = \mathbf{0}$ for all i, we must have the vector product $v \cdot (1, y(1), \ldots, y(t)) = \mathbf{0}$ which leads a contradiction. So we have proved the following result.

Now we prove the converse statement. Since $\mathbf{e}_0$ is linearly independent with $\mathcal{V}_{eq}$, the number of solutions of $(y(1), \ldots, y(t))$ satisfying N1 is exactly $2^{n(s-a)}$. Similarly, as every vector of $\mathcal{V}^*_{neq}$ is linearly independent with $\mathcal{V}_{eq}$ the number of solutions of $(y(1), \ldots, y(t))$ satisfying N1 and $v_0 + \sum_{j=1}^{t} v_j \cdot y(j) = \mathbf{0}$ for any fixed $(v_1, v_1, \ldots, v_t) \in \mathcal{V}^*_{neq}$ is exactly $2^{n(s-a-1)}$. These above statements follows from a simple facts from linea algebra which counts the number of solutions of a system of linear equations. So the number of solutions of y satisfying N1 and N2 is at least $2^{n(s-a)} \times (1 - \frac{\#\mathcal{V}^*_{neq}}{2^n})$ and every such y must be an output function with $\sim^y = \sim$ and $\#\mathbb{P}_n[y] = (2^n - s)!$ (lemma 4). So we obtain the lower bound of $\#\mathbb{P}_n[\sim]$. The upper bound follows from N1 and the fact that $\sim^y = \sim$ (the function y is determined by some specific $(s-a)$ distinct outputs of y due to the condition N1). □

4.3 Generator and number of accidents of a Collision Relation

So far we have defined input, output function and collision relation associated with any permutation π. Now we see that not all collisions are unexpected. There is a set of collision pairs which imply all other collisions independent of the permutation. Generator is the set representing the minimum such set and the number of such collisions are called accident.

Definition 6. *The generator* $\mathrm{Gen} := \mathrm{Gen}(\sim) = ((i_1, j_1), \ldots, (i_a, j_a))$ *of a relation* $\sim$ *corresponds to a maximal linearly independent set of vectors (also known as basis)* $\mathcal{B} := \{\mathbf{A}_{i_1} - \mathbf{A}_{j_1}, \ldots, \mathbf{A}_{i_a} - \mathbf{A}_{j_a}\}$ *of* $\mathcal{V}_{eq}$ *where the pairs of indices* (i_k, j_k)*'s are chosen as smallest as possible w.r.t. the dictionary order* $\prec$ *on* $[1,t]^{(2)} := \{(i,j) : i > j\}$.[6] *The number of accident is defined as* $a = \mathsf{acc}(\sim) := \mathrm{rank}(\mathcal{V}_{eq})$.

Note that the number of accidents a, and the generator Gen defined above must be unique which can be defined recursively as follows. The pair (i_k, j_k) is the smallest related pair (i,j) larger than (i_{k-1}, j_{k-1}) such that $(\mathbf{A}^{\sim}_i - \mathbf{A}^{\sim}_j)$ is not linearly independent with $\{\mathbf{A}^{\sim}_{i_c} - \mathbf{A}^{\sim}_{j_c} : c < k\}$. So a relation $\sim$ uniquely determines the generator $\mathrm{Gen}(\sim)$. Now we show that the converse is also true, i.e. a generator uniquely determines a relation.

Lemma 6. *Any relation* $\sim$ *satisfying the necessary condition of lemma 5 is uniquely determined by its generator* $\mathrm{Gen}(\sim)$. *Hence the number of collision relations with a accident is at most* $\binom{t}{2}^a$.

Proof. Suppose two relations $\sim$ and $\sim'$ have the same generator $\mathcal{B} = ((i_1, j_1), \ldots, (i_a, j_a))$ and they first time (w.r.t. dictionary order) differ on (i,j). We may w.l.o.g. assume that $i \sim j$ and $i \not\sim' j$. If $(i_{k-1}, j_{k-1}) \prec (i,j) \prec (i_k, j_k)$ then

[6] $(i,j) \prec (i',j')$ if and only if either $i < i'$ or $i = i', j < j'$. The notation $(i,j) \preceq (i',j')$ means either $(i,j) \prec (i',j')$ or $(i,j) = (i',j')$. Whenever we denote $i \sim j$ we mean $i > j$, i.e. $(i,j) \in [1,t]^{(2)}$. The notion of smaller and larger for pairs are based on the dictionary order.

$\mathbf{A}_i^{\sim} - \mathbf{A}_j^{\sim}$ is linearly dependent with $\{\mathbf{A}_{i_c}^{\sim} - \mathbf{A}_{j_c}^{\sim} : c \leq k-1\}$, o.w. (i,j) should be in the generator for $\sim$. As $\sim_{i-1}$=$\sim'_{i-1}$ and $\mathbf{A}$ is a lower triangular matrix we have $\mathbf{A}_{i'}^{\sim} = \mathbf{A}_{i'}^{\sim'}, \forall i' \leq i$. This implies that $\mathbf{A}_i^{\sim'} - \mathbf{A}_j^{\sim'}$ must be linearly dependent with $\{\mathbf{A}_{i_c}^{\sim'} - \mathbf{A}_{j_c}^{\sim'} : c \leq k-1\}$ which violates that $\sim'$ satisfies the necessary condition of a collision relation. The last part of the lemma is trivial as we can choose every pair of the generator in $\binom{t}{2}$ ways. □

Corollary 1. *For any collision relation of accident a, $\mathsf{Pr}[\sim_\Pi = \sim] \leq 1/\mathbf{P}(2^n - s + a, a)$. Hence $\mathsf{Pr}[\mathsf{acc}(\sim_\Pi) = a] \leq \frac{\binom{t}{2}^a}{\mathbf{P}(2^n - s + a, a)}$ and if $t < 2^{n/2-1}$ then $\mathsf{Pr}[\mathsf{acc}(\sim_\Pi) \geq 2] \leq \frac{t^2}{2^n}$.*

Proof. The above corollary follows from the estimate of the number of collision relations of accident a (lemma 6) and the probability for a collision relation which can be derived from lemma 5. The last part follows from the infinite sum. We leave readers to verify it. □

Remark 1. The generator of a collision relation actually represents the set of all unexpected collisions. Each unexpected collision can occur with probability roughly about $1/2^n$ and these are independent to each other (the corollary 1). All other collisions present in the collision relation are implied from these and the choices of messages. Note that CBC can have few initial collisions for two messages if the messages have common prefix. However, after when they differ for the first time, all collisions other than unexpected are implied from the unexpected collisions.

5 Examples of SADE

Now we show that all members of $\mathcal{C}$ are SADEs.

Theorem 2. *Member of $\mathcal{C}$ and (modified) non-singular DAG-based domain extensions are SADE.*

Before we prove it we first introduce forced relation. An equivalence relation $\sim^*$ is called **forced relation** if $\mathcal{V}_{\text{eq}} = \{\mathbf{0}^t\}$ (the singleton set containing zero vector) and $\mathcal{V}^*_{neq}$ does not contain the zero vector. So a forced relation clearly satisfies the necessary condition of collision relation and it would satisfy the sufficient condition if we assume that $t(t-1) < 2^n$.

Lemma 7. Uniqueness of the Forced Relation $\sim^*$
There exists one and only one forced equivalence relation $\sim^*$. Moreover, it is the only equivalence relation with zero accident satisfying the necessary condition (stated in lemma 5) for a collision relation. The forced relation is a collision relation if $t(t-1) < 2^n$.

Proof. The forced relation can be defined recursively. After we have defined the relation $\sim^*_{i-1}$ restricted on $[1, i-1]$, we extend the definition of $\sim^*_{i-1}$ to

$\sim_i^*$ restricted on $[1,i]$ as follows: $i \sim_i^* j$ if and only if $\mathbf{A}_i^{\sim_{i-1}^*} = \mathbf{A}_j^{\sim_{i-1}^*}$, $\forall j \leq i$. This defines an equivalence relation $\sim^*$ and it is easy to see that it satisfies the property that $\mathbf{A}_i^{\sim^*} = \mathbf{A}_j^{\sim^*}$ if and only if $i \sim^* j$ and hence $\sim^*$ is a forced relation. Clearly the $\mathrm{rank}(\mathcal{V}_{eq}) = 0$ and hence it is the only relation with accident zero. The forced relation is a collision relation whenever $t(t-1) < 2^n$ since $\#\mathcal{V}_{neq}^* < t(t-1)$. □

Lemma 8. *The forced relation is a sub-relation of all collision relations. In other words, if $i \sim^* j$ then $y^\pi(i) = y^\pi(j)$ for all permutation π. If $t(t-1) < 2^n$ then for any $i \not\sim^* j$ there exists a permutation π such that $y^\pi(i) \neq y^\pi(j)$.*

Proof. We prove that $i \sim^* j$, $j < i$ then $i \sim j$ for all collision relation $\sim$. We prove it by induction on i for any fixed collision relation $\sim$. Suppose the result is true up to $i-1$ and $i \sim^* j$. Hence $\sim_{i-1}^*$ is a sub-relation of $\sim$. So $\mathbf{A}_i^{\sim_{i-1}^*} = \mathbf{A}_j^{\sim_{i-1}^*}$ implies that $\mathbf{A}_i^{\sim_{i-1}} = \mathbf{A}_j^{\sim_{i-1}}$. From the previous lemma we know that $i \sim_i^* j$ if and only if $\mathbf{A}_i^{\sim_{i-1}^*} = \mathbf{A}_j^{\sim_{i-1}^*}$, $\forall j \leq i$ and hence $i \sim j$ (since zero vector can not be in $\mathcal{V}_{neq}^*$). The last part is obvious from that fact that the force relation is collision relation whenever $t(t-1) < 2^n$. □

Let $\mathcal{M}$ be a message space such that $\max_{M \in \mathcal{M}} \ell(m) < 2^{n/2-1}$. Then for any pair of messages the joint coefficient matrix has at most t rows such that $t(t-1) < 2^n$. So we can provide an equivalent definition of SADE (using lemma 8) using that we prove that every member of $\mathcal{C}$ is SADE:

Lemma 9. Equivalence characterization of *SADE*
A affine domain extension is SADE if and only if for any tuple of two distinct messages $\mathbf{M} = (M, M')$, the forced collision relation is I-isolated w.r.t. the joint coefficient matrix $\mathbf{A}^{\mathbf{M}}$ where $I = \{t := \ell, t' := \ell + \ell'\}$ corresponds to the final-index set.

Proof of Theorem 2. Lemma 1 shows that non-singular DAG-based constructions are SADE. We prove the result for CBC-MAC with prefix-free message space. The similar argument will work for other members of $\mathcal{C}$. Let $M = (\alpha_1, \cdots, \alpha_\ell)$ and $M' = (\alpha'_1, \cdots, \alpha'_{\ell'})$ be two prefix-free messages, i.e. one is not prefix to other. Suppose $s \geq 0$ with $\alpha_1 = \alpha'_1, \cdots, \alpha_s = \alpha'_s, \alpha_{s+1} \neq \alpha'_{s+1}$. Then $s < \min\{\ell, \ell'\}$ and it is called length of common prefix. Now define a collision relation $\sim$ such that $1 \sim \ell + 1, \cdots, s \sim s + \ell$ and all other unequal values are unrelated (clearly, $i \sim i$ for all i since it is an equivalence relation). Now let $\mathbf{A} = \mathbf{A}^{(M,M')}$ then it is easy to see that $\mathbf{A}_i^{\sim} = \mathbf{A}_j^{\sim}$ if and only if $i \sim j$. Hence it must be the trivial collision relation. Thus, CBC is SADE for any prefix-free message space. However, if we choose two messages such that one is prefix to other then clearly trivial collision relation says that CBC is not a secure affine domain extension. □

6 A Unified PRF Security Analysis for all Secure Affine Domain Extensions

A collision relation is called (t,q)-collision relation if it is induced by a matrix $\mathbf{A}_{t \times (t+1)}$ with a final index q-set I.

Definition 7. *Given a domain extension $\mathcal{D}$, let $N(t,q)$ denote the maximum number of (t,q)-collision relations with one accident such that I is not isolated where the maximum is taken over all joint coefficient matrices for coordinate-wise distinct q-tuple $\mathbf{M}=(M_1,\ldots,M_q)$ such that $\sum_{i=1}^{q}\ell(M_i)=t$.*

Clearly, $N(t,q)\leq \max_{\ell_1,\ldots,\ell_q}\sum_{j<i}N(\ell_i+\ell_j,2)$ where the maximum is taken over all q-tuples $(\ell_1,\ldots,\ell_q)$ such that the sum $\sum_i \ell_i \leq t$. So if $N(\ell_i+\ell_j,2)\leq c(\ell_i+\ell_j)$ for some constant c then $N(\ell,q)\leq ctq$. We use when we provide improved bounds for the members of $\mathcal{C}$.

Definition 8. *A permutation π is said to be* **w**-regular *if*

type-1: $R(y^\pi|_{\overline{I}})$ and $W=\{w_1,\ldots,w_q\}$ are disjoint and
type-2: $x^\pi(i)\neq x^\pi(j)$, for all $i\in I$ and $j\neq i$, i.e. $\sim^\pi$ is I-isolated.

Lemma 10. $\mathsf{Pr}_{\Pi\xleftarrow{*}\mathbb{P}_n}[y^\Pi(t_i)=w_i, 1\leq i\leq q|\Pi \text{ is } \mathbf{w}\text{-regular}]\geq \frac{1}{\mathbf{P}(2^n-1,q)}$

Proof. Note that whether a permutation is regular or not, can be determined from the intermediate output function $y^\pi|_{\overline{I}}$ (let us denote that $y^\pi|_{\overline{I}}\rightarrowtail x^\pi$). Let $V_{s,\text{reg}}$ be the set of all intermediate output functions g' with $\#R(g')=s, R(g')\cap W=\emptyset$ (type-1) and $f(i)\neq f(j)$, for all $i\in I$ and $j\neq i$ where $g'\rightarrowtail f:[1,t]\rightarrow\mathbb{F}_{2^n}$ (type-2). The set of all regular permutations can be partitioned into disjoint sets $\mathbb{P}_n[g']$ for all $g'\in V_{s,\text{reg}}$ and $1\leq s\leq t-q$. Each $g'\in V_{s,\text{reg}}$ can be extended to an output function g (i.e. $g|_{\overline{I}}=g'$) if and only if $g(I)\cap g(\overline{I})=\emptyset$ and hence there are exactly $\mathbf{P}(2^n-s,q)$ such output functions. Among these, there is only one output function g satisfying $g(t_i)=w_i, 1\leq i\leq q$. So $\mathsf{Pr}_{\Pi\xleftarrow{*}\mathbb{P}_n}[y^\Pi(t_i)=w_i, 1\leq i\leq q|\Pi\in\mathbb{P}_n[g']]\geq\frac{1}{\mathbf{P}(2^n-1,q)}$ for each function $g'\in V_{s,\text{reg}}$. Since $V_{s,\text{reg}}$'s are partitions of the set of all regular permutation $\square$

Lemma 11. *An estimate of probability of type-1 regular permutation:*

$$\mathsf{Pr}_{\Pi\xleftarrow{*}\mathbb{P}_n}[y^\Pi(i)\notin W,\ \forall i\in\overline{I}]\geq\frac{\mathbf{P}(2^n-q,t)}{\mathbf{P}(2^n,t)}\geq\big(1-\frac{q}{2^n-t}\big)^t. \tag{3}$$

Proof. Let $S_r=\{\pi: y^\pi(i)\notin W, 1\leq i\leq r\}$. We first prove that

$$\#S_r\geq\mathbf{P}(2^n-q,r)\times(2^n-r)!,\ 1\leq r\leq t. \tag{4}$$

The first inequality of eq. 3 follows from this by choosing $r=t$ in the above equation. The second inequality is straightforward from the relation $(2^n-q-i)/(2^n-i)\geq(1-q/(2^n-t))$ for all $i\leq t$.

Now we prove the eq. 4 by induction on r.

For $r=1$, we know that $x^\pi(1)=a_{1,0}$ and hence any permutation π such that $\pi(a_{1,0})\notin W$ belongs to S_1. Hence $\#S_1\geq(2^n-q)\times(2^n-1)!$ and so the statement is true for $r=1$.

By induction hypothesis, we assume the statement for r and we want to prove it for $r+1$, $1\leq r<t$. We can partition the set S_r into disjoint sets $S_{r,g}:=\{\pi: y^\pi(i)=g(i), 1\leq i\leq r\}$ over all choices of functions $g:[1,r]\rightarrow\mathbb{F}_{2^n}\setminus W$

(let us denote the set of all such functions by G). That is, $S_r = \bigsqcup_{g\in G} S_{r,g}$. Let $S'_{r+1,g} := \{\pi : y^\pi(i) = g(i), 1 \le i \le r, y^\pi(r+1) \notin W\}$ then $S_{r+1} = \bigsqcup_{g\in G} S'_{r+1,g}$. The main observation for the lemma is the following claim.

Claim: $\frac{\#S_{r+1,g}}{\#S(r,g)} \ge 1 - \frac{q}{2^n - r}$, $\forall g \in G$.

Let $r' = \#R(f) \le r$ and $g \rightarrowtail f$. If $(a_{j,0} + \sum_{i=1}^{r} a_{j,i} \cdot g(i)) \notin R(f)$ then $\#S_{r+1,f} \times (2^n - r') = \#S(r,f) \times (2^n - q - r')$. Otherwise, $\#S_{r+1,f} = \#S(r,f)$. So we have proved the claim. The induction statement follows immediately from this claim. □

If we denote the probability for type-2 regular permutations by $(1-\epsilon)$ for some non-negative ϵ then a randomly chosen permutation is regular has probability at least $(1 - q/(2^n - t))^t - \epsilon$ (applying union bound of probability theory). Hence by applying lemma 10 we have our one of the main results.

Proposition 1. *Let* $\mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[x^\Pi(t_i) = x^\Pi(j)$ *for some* $j \ne i] \le \epsilon$ *then* $\mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[y^\Pi(t_i) = w_i, 1 \le i \le q] \ge \frac{(1-q/(2^n-t))^t - \epsilon}{\mathbf{P}(2^n-1,q)}$. *Hence (applying decorrelation theorem) for any affine domain extension* $\mathcal{D}$ *we have*

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \le \frac{qt}{2^n - t} + \epsilon + \frac{q(q-1)}{2^{n+1}}$$

$$\le \frac{3qt}{2^n} + \epsilon \text{ if } t < 2^{n-1}$$

Due to the above theorem, given any domain extension $\mathcal{D}$ it remains only to bound the probability $\epsilon := \mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[x^\Pi(t_i) = x^\Pi(j)$ for some $j \ne i] = \mathsf{Pr}[\sim^\Pi$ is I-isolated].

Theorem 3. $\mathsf{Pr}[\sim^\Pi$ *is* I*-isolated*$] \le N(t,q)/2^n + \ell^4 q^2/2^{2n}$ *where* $\ell = \max_i \ell_i$. *Hence*

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \le \frac{3qt}{2^n} + N(t,q)/2^n, \quad \text{if } \ell < 2^{n/3-1}.$$

Proof. The probability that $\sim^\Pi$ has rank two or more is less than $\sum_{1\le j<j'\le q} \frac{(\ell_j + \ell_{j'})^4}{2^{2n}}$ (by using corollary1) which is less than $8\ell^3 tq/2^{2n}$ (a simple algebra) where $\ell = \max_i \ell_i$. So if $\ell < 2^{n/3-1}$ then $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \le \frac{3qt}{2^n} + N(t,q)/2^n$. □

Theorem 4. $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \le \frac{t^2}{2^n - 2}$.

Proof. The result is immediate from the theorem 3 and lemma 6. However we need the restriction on ℓ. If we want to prove unconditional bound then we use the lemma **??** on the forced collision relation $\sim^*$. Note that $\mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[x^\Pi(t_i) \ne x^\Pi(j)$ for all $j \ne i] \ge \mathsf{Pr}[\sim^\Pi = \sim^*] \ge (1 - t(t-1)/2^n)$ (lemma **??**) and hence $\epsilon \le t(t-1)/2^n$. Thus, by using proposition 1 we proved that $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \le \frac{t^2}{2^n-2}$. □

7 Improved security bound for $\mathcal{C}$

We provide a sketch (the detail can be found in the full version of the paper) of improved security analysis of members of $\mathcal{C}$. Because of theorem 3, we need to bound $N(t,q)$. Note that $N(t,q) \leq \sum_{i,j} N(\ell_i + \ell_j, 2)$. So we choose two messages $M \neq M'$, $\ell = \ell(M)$ and $\ell' = \ell(M')$ and want to bound $N(\ell + \ell', 2)$. More precisely, we bound the number of collision relation with one accident which is not t-isolated where $t = \ell + \ell'$. We use the following lemmas proved in [2] (lemma 12 and 17 of [2]).

Lemma 12. *Let $\mathcal{D}$ be the CBC domain extension. The number of collision relations with one accident so that $\ell \sim \ell + \ell'$ is bounded by $d'(|\ell - \ell'|)$ where $d(\ell)$ denote the number of divisors of ℓ and $d'(\ell)$ denote the maximum $d(\ell')$ over all choices of $\ell' \leq \ell$. (by convention, $d(0) = 1$. A trivial bound is $d'(\ell) \leq \ell$.*

Lemma 13. *Let $\mathcal{D}$ be the CBC domain extension. The number of t-isolated collision relations with one accident is bounded by $4(\ell + \ell')$.*

Improved Security Bound for CBC By applying the lemma 12 we have $N(t,q) \leq 8tq$. Hence the CBC-MAC for prefix-free message space has the following PRF advantage:

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{CBC}}(q,t,\ell) \leq \frac{11tq}{2^n} \text{ if } \ell \leq 2^{n/3-1}$$

Improved Security Bound for GCBC

If (i,j) is a basis of a collision relation $\sim$ with one accident where both $i,j \notin \{1,\ell,\ell+1,t\}$ then the basis vector $\mathbf{A}^{\sim}_i - \mathbf{A}^{\sim}_j = c \cdot \mathbf{e}_0 + \mathbf{e}_{i-1} + \mathbf{e}_{j-1}$ for some constant c. has only two non-zero entries which are $\mathbf{1}$ with the column index 1 or more (ignoring the zeroth column). Let j and j' denote the column index.
Case-A : $\delta_M \neq \delta_{M'}$: In this case one can show easily that $\sim$ is t-isolated. If not $t \sim k$ for some $k \neq t$ then $\mathbf{A}^{\sim}_k - \mathbf{A}^{\sim}_t$ can not be multiple of $c \cdot \mathbf{e}_0 + \mathbf{e}_{i-1} + \mathbf{e}_{j-1}$.
Case-B : $\delta_M = \delta_{M'}$ **and** $x_\ell \neq x_{\ell'}$: $\sim$ is t-isolated unless $\ell \sim t$. This implies either $\ell - 1$ or $t - 1$ is related to i ($< j$ say). Let $\ell - 1 \sim i - 1$. Now $\mathbf{A}^{\sim}_{i-1} - \mathbf{A}^{\sim}_{\ell-1}$ is multiple of $c \cdot \mathbf{e}_0 + \mathbf{e}_{i-1} + \mathbf{e}_{j-1}$. This is possible only if $\mathbf{A}^{\sim}_{i-1} = \mathbf{A}^{\sim}_{\ell-1}$ and hence $i - 2 \sim \ell - 2$ and so on. So we get $1 \sim \ell - i + 1$ which can not be true as $\mathbf{A}^{\sim}_1 - \mathbf{A}^{\sim}_{\ell-i+1}$ can not be multiple of $c \cdot \mathbf{e}_0 + \mathbf{e}_{i-1} + \mathbf{e}_{j-1}$. Similarly one can prove that when $i - 1 \sim t - 1$.
Case-C : $\delta_M = \delta_{M'}$ **and** $x_\ell = x_{\ell'}$: In this case we reduces to CBC case by dropping the last message block from both the message.

So we can assume that one of the i,j from the set $\{1,\ell,\ell+1,t\}$ and hence $N(2,t) \leq 8t$. Hence

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{GCBC}}(q,t,\ell) \leq \frac{11tq}{2^n} \text{ if } \ell \leq 2^{n/3-1}$$

Security Bound for OMAC

Case-A : $\delta_M \neq \delta_{M'}$: Suppose I is not isolated in a collision relation $\sim$ of rank one and say $t \sim i'$. Let $\{(i,j)\} = \mathbb{B}$ such that $i,j \notin I$ then the first element in $\mathbf{A}_t^{\sim} - \mathbf{A}_{i'}^{\sim}$ is non-zero (either $c_{\delta'} - c_\delta$ or $c_{\delta'} - 1$ or $c_{\delta'}$), whereas the first element of $\mathbf{A}_i^{\sim} - \mathbf{A}_j^{\sim}$ is zero. Thus, the rank should be more than one. Hence, the only possible collision relation of rank one are relations with the basis (i,j) where $j \in I$. So, the number of such relations is at most $2(\ell + \ell')$.

Case-B : $\delta_M = \delta_{M'}$ Suppose we have $t \sim i'$ where $i' \notin I$ then by similar reason, the basis should contain the pair whose one element is from I. So there are at most $2(\ell + \ell')$ many such relations. Now we consider the case when $\ell \sim t$. This implies that $\mathsf{CBC}(\overline{M}) = \mathsf{CBC}(\overline{M}')$ and accident is still one for CBC. Since $\delta_M = \delta_{M'}$, $\overline{M} \neq \overline{M}'$. Now by using lemma 12 (also in Lemma 12 of [2]), we know that there are at most $d(|\ell - \ell'|)$ such relations with one accident.

Combining the above two cases, the total number of collision relations with one accident is at most $3(\ell + \ell')$ and hence $N(t,q) \leq 6tq$. Thus, we have PRF-insecurity bound for OMAC as

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathrm{OMAC}} \leq \frac{9qt}{2^n} \text{ if } \ell \leq 2^{n/3-1}$$

Security Bound for PMAC It is easy to see that basis of an accident one collision relation must contain a final index since the first column entry of row ℓ or t is c_δ which is different from those of all other rows. So $N(\ell+\ell', 2) \leq 2(\ell+\ell')$ and hence $N(t,q) \leq 2tq$. So, PMAC has the following security bound.

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathrm{PMAC}} \frac{5qt}{2^n} \text{ if } \ell \leq 2^{n/3-1}$$

Theorem 5. *Each member of $\mathcal{C}$ has PRF advantage $O(tq/2^n)$ if $\ell < 2^{n/3-1}$.*

8 Conclusion and Future Work

We provide an unified framework for improving PRF advantages of many known blockcipher based domain extensions. We obtain improved bounds $O(tq/2^n)$ for all members of $\mathcal{C}$ and our general result can also help to obtain similar improved bound for any affine domain extension, once we know a better estimate of $N(t,q)$. We believe that $N(t,q) = O(tq)$ for all secure affine domain extension and this would be an interesting research area to prove it. The other possible direction of research is to go further beyond $O(tq/2^n)$. To do so we need to find a completely new proof technique as our general bound or others proof idea for improved bounds can not do so.

References

1. Mihir Bellare and Roch Guérin and Phillip Rogaway, XOR MACs: New Methods for Message Authentication Using Finite Pseudorandom Functions, CRYPTO 1995, Volume **963**, pp 15-28.

2. M. Bellare, K. Pietrzak and P. Rogaway. Improved Security Analysis for CBC MACs. Advances in Cryptology - CRYPTO 2005. Lecture Notes in Computer Science, Volume **3621**, pp 527-545.

3. M. Bellare, J. Killan and P. Rogaway. The security of the cipher block chanining Message Authentication Code. Advances in Cryptology - CRYPTO 1994. Lecture Notes in Computer Science, Volume **839**, pp 341-358.

4. Daniel J. Bernstein. A short proof of the unpredictability of cipher block chaining (2005). URL: http://cr.yp.to/papers.html#easycbc. ID 24120a1f8b92722b5e1 5fbb6a86521a0.

5. J. Black and P. Rogaway. CBC MACs for arbitrary length messages. Advances in Cryptology - CRYPTO 2000. Lecture Notes in Computer Science, Volume **1880**, pp 197-215.

6. J. Black and P. Rogaway. A Block-Cipher Mode of Operations for Parallelizable Message Authentication. Advances in Cryptology - Eurocrypt 2002. Lecture Notes in Computer Science, Volume **2332**, pp 384-397.

7. I. B. Damgård. *A Design Principle for Hash Functions*. Advances in Cryptology - Crypto'89, Lecture Notes in Computer Sciences, vol 435, Springer-Verlag, pp. 416-427, 1989.

8. Oded Goldreich, Shafi Goldwasser and Silvio Micali, How to construct random functions, JACM 1986, Volume **33-4**, pp 792-807.

9. T. Iwata and K. Kurosawa. OMAC : One-Key CBC MAC. Fast Software Encryption, 10th International Workshop, FSE 2003. Lecture Notes in Computer Science, Volume **2887**, pp 129-153.

10. T. Iwata and K. Kurosawa. Stronger Security Bounds for OMAC, TMAC, and XCBC. Progress in Cryptology - INDOCRYPT 2003. Lecture Notes in Computer Science, Volume **2904**, pp 402-415.

11. C. S. Jutla. PRF Domain Extension using DAG. Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006. Lecture Notes in Computer Science, Volume **3876** pp 561-580.

12. K. Kurosawa and T. Iwata. TMAC : Two-Key CBC MAC. Topics in Cryptology - CT-RSA 2003: The Cryptographers' Track at the RSA Conference 2003. Lecture Notes in Computer Science, Volume **2612**, pp 33-49.

13. M. Luby and C. Rackoff, How to construct pseudo-random permutations from pseudo-random functions, Advances in Cryptology, Crypto 1985, Lecture Notes in Computer Science 1984, Volume **218**, Springer-Verlag, pp 447.

14. K. Minematsu and T. Matsushima Improved Security Bounds for PMAC, TMAC, and XCBC. Fast Software Encryption 2007.

15. M. Nandi and A. Mandal Improved Security Analysis of PMAC. Journal of Mathematical Cryptology, July 2008, Volume **2**, No. 2 : pp 149162.

16. R. Merkle. *One Way Hash Functions and DES*. Advances in Cryptology - Crypto'89, Lecture Notes in Computer Sciences, Vol. 435, Springer-Verlag, pp. 428-446, 1989.

17. Mridul Nandi, Improved security analysis for OMAC as a pseudorandom function. Journal of Mathematical Cryptology. Volume **3**, Issue 2, pp 133148, 2009.

18. Mridul Nandi, Fast and Secure CBC-Type MAC Algorithms, FSE 2009, Lecture Notes in Computer Science, Volume **5665**, pp 375-393.

19. M. Nandi A Simple and Unified Method of Proving Indistinguishability. Progress in Cryptology - INDOCRYPT 2006. Lecture Notes in Computer Science, Volume **4329**, pp 317-334.

20. J. Patarin, Etude des Générateurs de Permutations Basés sur le Schéma du D.E.S., *Phd Thèsis de Doctorat de l'Université de Paris 6*, 1991.

21. E. Petrank and C. Rackoff. CBC MAC for real-time data sources. Journal of Cryptology, vol. 13, no. 3, pp. 315338, 2000.

22. Krzysztof Pietrzak, A Tight Bound for EMAC. ICALP (2), 2006, pp 168-179.

23. Palash Sarkar. Pseudo-Random Functions and Parallelizable Modes of Operations of a Block Cipher, Available in http://eprint.iacr.org/2009/217

24. S. Vaudenay. Decorrelation over infinite domains: the encrypted CBC-MAC case. Communications in Information and Systems (CIS), Volume 1, pp. 7585, 2001.

25. Serge Vaudenay, Decorrelation: A Theory for Block Cipher Security, J. Cryptology, Volume **16**, no 4, 2003, pp 249-286.

Appendix

Proof of Decorrelation Theorem

W.l.o.g we consider deterministic distinguisher $\mathcal{A}$ and the queries to be distinct. So the final output of $\mathcal{A}$ only depends on responses $w_1, \ldots, w_q$. Let $S \subseteq \mathbb{F}_{2^n}^q$ be the set of all possible q-tuple of responses on which $\mathcal{A}$ returns 1. Now for any fixed $\mathbf{w} := (w_1, \ldots, w_q)$, let $\mathbf{M} := \mathbf{M}(\mathbf{w}) = (M_1, \ldots, M_q)$ be the corresponding distinct queries. Note that these queries are fixed and independent of oracles. Let $\mathcal{Y}$ be the set of all coordinate-wise distinct elements from $\mathbb{F}_{2^n}^q$. So $\mathsf{Pr}[\mathcal{A}^{\mathcal{D}^{\Pi}} = 1 : \Pi \xleftarrow{*} \mathbb{P}_n] = \sum_{\mathbf{w} \in S} \mu_{\mathbf{w},\mathbf{M}(\mathbf{w})}$. Let $\mathbf{M}(\mathbf{w}) = (M_1, \ldots, M_q)$.

$$
\begin{aligned}
\mathbf{Adv}_{\mathcal{D}}^{\mathrm{prf}}(\mathcal{A}) &= \frac{\#S}{2^{nq}} - \sum_{\mathbf{w} \in S} \mathsf{Pr}[\mathcal{D}^{\Pi}(M_1) = w_1, \ldots, \mathcal{D}^{\Pi}(M_q) = w_q : \Pi \xleftarrow{*} \mathbb{P}_n] \\
&\leq \frac{\#\mathcal{S} \setminus \mathcal{Y}}{2^{nq}} + \sum_{\mathbf{w} \in S \cap \mathcal{Y}} \big(\frac{1}{2^{nq}} - \mathsf{Pr}[\mathcal{D}^{\Pi}(M_1) = w_1, \ldots, \mathcal{D}^{\Pi}(M_q) = w_q : \Pi \xleftarrow{*} \mathbb{P}_n] \big) \\
&\leq \frac{q(q-1)}{2^{n+1}} + \frac{\epsilon \times \#(S \cap \mathcal{Y})}{2^{-nq}} \quad \text{(from the given condition of the theorem)} \\
&\leq q(q-1)/2^{n+1} + \epsilon \qquad \square
\end{aligned}
$$