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

Mridul Nandi

National Institute of Standards and Technology and
The George Washington University, Computer Science Department
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, a class of ADEs (affine domain extensions) and a wide subclass of SADEs (secure ADEs) are introduced in the paper which contain following constructions $\mathcal{C} = \{$CBC-MAC, GCBC*, OMAC, PMAC$\}$. We prove 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 and $N(t,q)$ is a parameter defined in the paper. The PRF advantage of any SADE is $O(t^2/2^n)$ as we can show that $N(t,q) \leq \binom{t}{2}$. Moreover, $N(t,q) = O(tq)$ for all members of $\mathcal{C}$ and hence these MACs have *improved advantages* $O(tq/2^n)$. Eventually, our proposed bounds for CBC-MAC and GCBC* become strictly 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 $(\mathbb{F}_{2^n},+,\cdot,\mathbf{0},\mathbf{1})$ treated equally in the paper).

1.1 Related Works: PRF Security Analysis of Known MACs

The 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* [19][1], PMAC [6]$\}$ and (directed acyclic graph) DAG-based PRFs [11, 20]. 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 (Jutla's class) was proposed where each *non-singular* DAG or a family of non-singular DAGs (see Definition 4) corresponds to a domain extension. 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. Even though the Jutla's class contains CBC, GCBC* and others, it does not include those which encrypt a constant block (e.g. OMAC and PMAC encrypt the zero block $\mathbf{0}$). If we add a special node representing to the encryption of $\mathbf{0}$ then OMAC and PMAC can be included (as described in Nandi's class [20]).

(2)Finding Improved Bounds of PRF Advantages: The original PRF bound for the members of $\mathcal{C}$ and DAG-based constructions is about $t^2/2^n$ (sometimes $\ell^2q^2/2^n$) [3–6, 9–11, 20, 22]. The improved bound $\ell q^2/2^n$ for CBC-MAC was shown first time in [2]. Afterwards, similar or better improved bounds were shown for PMAC, OMAC [14, 15, 18] and others, e.g. EMAC [22, 23], XCBC and TMAC [5, 12, 14] (see Table 1 for existing PRF bounds).

1.2 Motivation and Our Results

(1) We Unify Many Known Domain Extensions. It is always worthwhile to unify similar objects and study them under one umbrella. It helps us to understand the basic proof nature and to come up with new efficient constructions. 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 via some affine functions. Moreover, the output of the domain extension is the last output of π.

[1] GCBC* is one example of one-key GCBC [19], 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].

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 $\mathbf{C} = ((c_{i,j}))_{1\le i\le \ell}^{0\le j\le \ell}$ (i.e. $c_{i,j} = \mathbf{0}$, for all $i \le j$) and $\ell = \ell(M)$ such that*

$$\mathcal{D}^{\pi}(M) = y(\ell),\ \forall \pi \in \mathbb{P}_n \text{ (the set of all permutations on } \mathbb{F}_{2^n})$$

where the i^{th} intermediate output $y(i) = \pi(x(i))$ and the i^{th} intermediate input $x(i) = c_{i,0} + \sum_{j=1}^{i-1} c_{i,j} \cdot y(j),\ 1 \le i \le \ell$. The matrix $\mathbf{C}$ is called coefficient matrix *corresponding to M.*

In Section 4, we identify a class of PRF secure domain extensions and call them SADE (*secure* affine domain extensions). It contains all modified non-singular DAG-based PRFs (i.e. Nandi's class) and the members of $\mathcal{C}$. In Theorem 2 we prove that CBC-MAC with prefix-free message space, OMAC, PAMC and GCBC* are secure affine domain extensions. Non-secure ADE does not necessarily mean insecure constructions. 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 keys (auxiliary keys or/and blockcipher keys). They mostly have underlying CBC type structure. For example, the PRF security of EMAC can be reduced to collision probability of a ADE CBC-MAC [23]. Generalized CBC class or GCBC includes these constructions and they have been studied in [19]. This is beyond scope of our paper.

(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 $t^2/2^n$). So we need to find a different method to obtain improved bounds. The proof idea of [2] for the CBC-MAC uses structure graph which is also not suitable for a general ADE.

We use equivalence relation which is more appropriate to describe collision patterns on inputs of the blockcipher π during the computation of a ADE outputs. Given any q messages $M_1, \ldots, M_q$, let $t = \sum_{i=1}^{q} \ell_i$, $\ell_i = \ell(M_i)$ and $t_i = \sum_{j=1}^{i} \ell_i$. We write all intermediate inputs $x(1), \ldots, x(t)$ (also outputs $y(1), \ldots, y(t)$) in the order of the computations of $\mathcal{D}^{\pi}(M_1), \ldots, \mathcal{D}^{\pi}(M_q)$. We call the function x^{π} or $x : [1, t] \to \mathbb{F}_{2^n}$ *intermediate input function* associated with permutation π (and the messages $M_1, \ldots, M_q$). Similarly we define *intermediate output function* y (or y^{π}). Since intermediate inputs are eventually affine functions of intermediate outputs, there is a $t \times (t+1)$ lower triangular matrix $\mathbf{A}$ (called *joint coefficient matrix*) such that $\mathbf{A} \cdot \overline{\mathbf{y}} = \mathbf{x}$ where $\overline{\mathbf{y}} = (\mathbf{1}, y(1), \ldots, y(t))$ and $\mathbf{x} = (x(1), \ldots, x(t))$, called intermediate output and input vectors respectively.

Definition 2 (Collision Relation). *To any permutation π and a tuple of q messages $(M_1, \ldots, M_q) \in \mathcal{M}^q$ we associate an equivalence relation $\sim$ (called* collision relation*) on $[1, t] := \{1, 2, \ldots, t\}$ where all colliding inputs of π during the*

computations of $\mathcal{D}^\pi(M_1), \ldots, \mathcal{D}^\pi(M_q)$ represents the equivalence classes. More precisely, $i \sim j$ if and only if $x(i) = x(j)$ where $x(1), \ldots, x(t)$ is the sequence of all inputs of π while computing outputs of $M_1, \ldots, M_q$.

Whenever we fix messages $M_1, \ldots, M_q$ or understood from the context, we denote the collision relation by $\sim^\pi$. The collision relation characterizes that which intermediate inputs collide and which do not, independent of the actual values of inputs. There may be more than one permutation associating a collision relation. On the other hand, there may exist equivalence relations which are not collision relations for any permutation. We provide a characterization of collision relation in Lemma 2. We later see that some collisions on inputs (i.e. $x(i) = x(j)$ for $i \neq j$) are trivially derived from the previous occurred collisions (see Example 1). A set of accidents is the largest set of collisions which can not be derived from the other collisions only. All non-accident collisions are derived form the accidents. The set of accidents is very much analogous with a basis of a vector space[2]. A more formal definition of accidents in terms of basis of a vector space is given in Definition 3. Let

$$N(t,q) = \max_{\substack{(q,t)-\text{messages} \\ (M_1,\ldots,M_q)}} \sum_{1 \le i < i' \le q} N(M_i, M_{i'})$$

where $N(M_i, M_{i'})$ is the number of collision relations $\sim$ with one accident such that $\mathcal{D}^\pi(M_i)$ or $\mathcal{D}^\pi(M_{i'})$ collide with a different intermediate output. A precise definition is given in Section 6. In Theorem 3 we prove that for any SADE $\mathcal{D}$, $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \le \frac{3qt+N(t,q)}{2^n}$ if $\ell \le 2^{n/3-1}$. Moreover, in Theorem 4 we show that $N(t,q) \le \binom{t}{2}$ and hence $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \le \frac{t^2}{2^{n-2}}$.

(3)We Find Improved PRF Bounds for CBC and GCBC*. Because of Theorem 3, we only need to have a better estimation of $N(t,q)$ for a given SADE. In Section 6, 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 $\mathcal{C}$. In Theorem 5 we show that CBC with prefix-free message space, OMAC, GCBC*, PMAC have PRF advantages $O(tq/2^n)$.

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*[19]. Note that the bound $\ell q^2/2^n$ can be worse compare to $tq/2^n$ 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.

[2] In fact, it is defined via a basis or *generator* of a set of vectors $\mathcal{V}_{eq}$ defined in Section 3.1 (also see Section 3.2).

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

Table 1. PRF bounds for (q, t, ℓ)-distinguishers. R1: $\ell < 2^{n/3-1}$.

2 Preliminaries

We follow the notations described in introduction throughout the paper. We write $\mathbf{P}(m, r) = m(m-1)\cdots(m-r+1)$. We denote $(i,j)^{\text{th}}$ entry and i^{th} row of an $s \times (s+1)$-matrix $\mathbf{A}$ by $a_{i,j}$ and $\mathbf{A}_i$ 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'. We denote the functions having domain $[1, t] := \{1, 2, \ldots, t\}$ (index-set) by x, y, f, g etc.

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 $Ld(\sim) := \{1_1, \ldots, 1_s\}$. 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 $g(i) = g(j)$.[3] Two function f and g of same domain are said to be equality-matching if $\sim^f = \sim^g$ and we denote it by $f \stackrel{\circ}{=} g$.

2.1 Decorrelation Theorem

We state a useful result (Lemma 22 of [26]) for PRF security analysis (the result is also applicable for (strong) pseudorandom permutation [24], pseudo online cipher [20], etc.) and we call it *Decorrelation Theorem*. The main idea of the theorem was described as Patarin's "coefficient H-techniques" [21] (according to Vaudenay [26, 25]). Different generalized versions of the theorem are stated in [4, 20]. For a deterministic adaptive distinguisher, the queries and 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 $\frac{1}{2^{nq}}$ (equals to q-decorrelation probability for RF on $\mathbb{F}_{2^n}$) 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. A

[3] We distinguish the notation $\sim^g$ (induced relation from a function) and $\sim^\pi$ (collision relation associated with a permutation π) as g is a function and π is a permutation.

big advantage in the computation of $\mu_{\mathbf{M},\mathbf{w}}$ is that the source of randomness is only from the uniform distribution of Π over $\mathbb{P}_n$ which we denote by $\Pi \xleftarrow{*} \mathbb{P}_n$. We write the set $\{w_1, \ldots, w_q\}$ by W. 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 the maximum is taken over all (q, t, ℓ)-distinguishers.

Theorem 1. Decorrelation Theorem (Lemma 22 of [26])
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 $\sum_{i=1}^{q} \ell(M_i) \leq t$ and $\max_i \ell(M_i) \leq \ell$. Then $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}^{\Pi}}(q, t, \ell) \leq \epsilon + \frac{q(q-1)}{2^{n+1}}$.

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. The proof of the above theorem is given in the 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 Results on Intermediate Input, Output functions and Collision Relations

From now onwards we fix an affine domain extension $\mathcal{D}$ (see Definition 1) and q distinct messages $M_1, \ldots, M_q$, $q \geq 1$. Let $\mathbf{A}_{t \times (t+1)} = ((a_{i,j}))$ be its joint coefficient matrix (see Section 1). Note that when $q = 1$ the joint coefficient matrix is nothing but the coefficient matrix. We denote the i^{th} row of the matrix by $\mathbf{A}_i$. Recall that to any permutation π, we associate an intermediate input function x (or x^{π}) and output function y (or y^{π}), respectively where $x(i) = \sum_{j=0}^{i-1} a_{i,j} \cdot y(j)$ (denoted by $y \rightarrowtail x$) and $\pi(x(i)) = y(i), \forall i$.

3.1 Intermediate Output Function and Collision Relation

An intermediate output function y can be associated with more than one permutations, e.g. $y^{\pi_1} = \cdots = y^{\pi_r}$ for some permutations $\pi_1, \ldots, \pi_r$. Let $\mathbb{P}_n[y]$ denote the set of all permutations with y as an output function. In other words, $\mathbb{P}_n[y]$ is the collection of all permutations so that the computation of $\mathcal{D}^{\pi}(M_1), \ldots, \mathcal{D}^{\pi}(M_q)$ gives exactly the same sequence of intermediate inputs and outputs namely $x(1)$,

$y(1)$, ..., $x(t)$, $y(t)$. Clearly, all these permutations have to agree on the sets of all intermediate inputs as $\pi(x(i)) = y(i), \forall i, \forall \pi \in \mathbb{P}_n[y]$. We denote the number of distinct elements of $x(i)$'s (or $y(i)$'s) by s. Hence $\#\mathbb{P}_n[y] = (2^n - s)!$ whenever y is an output function. Recall that $x \stackrel{\circ}{=} y$ if $x(i) = x(j) \Leftrightarrow y(i) = y(j)$ (i.e. the collision patterns of x and y are the same) which is a necessary condition for intermediate input and output functions. So a function $y : [1, t] \rightarrow \mathbb{F}_{2^n}$ is not an intermediate output function if $x \not\stackrel{\circ}{=} y$ where $y \rightarrowtail x$.

Lemma 1. (characterization of an intermediate output function). *A function $y : [1, t] \rightarrow \mathbb{F}_{2^n}$ is an intermediate output function if and only if $y \stackrel{\circ}{=} x$ where $y \rightarrowtail x$ and in this case $\#\mathbb{P}_n[y] = (2^n - s)!$.*

We recall that a collision relation is the equivalence relation capturing the collision pattern of x^π (and equivalently y^π) associated with π (see Definition 2). We have already characterized intermediate output function in Lemma 1. In this section we provide a characterization of collision relations since all equivalence relations on $[1, t]$ are not necessarily 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 $\mathbf{v}^\sim = (v_0, v_1^\sim, \ldots, v_t^\sim)$ where $v_i^\sim = \sum_{j \in [[i]]} v_j$, if $i \in Ld(\sim)$, o.w. $v_i^\sim = \mathbf{0}$. We mainly consider the $\sim$-reduced vectors for all row vectors $\mathbf{A}_i$'s of the joint coefficient matrix $\mathbf{A}$. If $\overline{\mathbf{y}} = (\mathbf{1}, y(1), \ldots, y(t))$ for an intermediate output function inducing a collision relation $\sim = \sim^y$ then $(\mathbf{A}_i^\sim - \mathbf{A}_j^\sim) \cdot \overline{\mathbf{y}} = x(i) - x(j) = \mathbf{0}$ if and only if $i \sim j$. So we also define the following sets of $(t+1)$-vectors.

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

Thus, $\forall \mathbf{v} \in \mathcal{V}_{eq}$, $\mathbf{v} \cdot \overline{\mathbf{y}} = \mathbf{0}$. Similarly, $\forall \mathbf{v} \in \mathcal{V}_{neq}$, $\mathbf{v} \cdot \overline{\mathbf{y}} \neq \mathbf{0}$. Let $\mathbf{e}_k \in \mathbb{F}_{2^n}^{t+1}$ be the $(t+1)$-vector whose k^{th} entry is $\mathbf{1}$ and all others are $\mathbf{0}$, $0 \leq k \leq t$. Then $(\mathbf{e}_i - \mathbf{e}_j) \cdot \overline{\mathbf{y}} = y(i) - y(j) \neq \mathbf{0}$ whenever $i \not\sim j$. Also, $\mathbf{e}_0 \cdot \overline{\mathbf{y}} = \mathbf{1} \neq \mathbf{0}$. So we define the following two sets of vectors:

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

We have $\forall \mathbf{v} \in \mathcal{V}_{neq}^{**}$, $\mathbf{v} \cdot \overline{\mathbf{y}} \neq \mathbf{0}$. In summary, intermediate output vector $\overline{\mathbf{y}}$ is in the null space of the set of vectors $\mathcal{V}_{eq}$ and not in the null space of $\mathcal{V}_{neq}^{**}$. So clearly a necessary condition for a collision relation is that the vectors of $\mathcal{V}_{neq}^{**}$ is not in the span of $\mathcal{V}_{eq}$. This necessary condition is also a sufficient condition as described in the following lemma. We provide an estimate of the size of the set $\mathbb{P}_n[\sim] = \{\pi : \sim^\pi = \sim\}$, the set of all permutations associating the collision relation $\sim$ (a similar definition is given for $\mathbb{P}_n[y]$ and we distinguish this two notation by the argument y which is a function, and $\sim$ which is an equivalence relation). We use the well known result regarding the number of solutions of a system of linear equations.

Lemma 2 (Characterization of Collision Relation). *Let $\sim$ be a collision relation then there exists $y : [1, t] \rightarrow \mathbb{F}_{2^n}$ 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* $a = \mathsf{acc}(\sim) := \mathrm{rank}(\mathcal{V}_{eq})$ *(the number of accidents). Hence* $\sim$ *is a collision relation if* $\#\mathcal{V}^{*}_{neq} < 2^n$.

The proof of the lemma is given in the full version of the paper [17]. We provide a sketch of the proof. It is easy to see that if both N1 and N2 are true for some function y then $y \stackrel{\circ}{=} x$ for $y \rightarrowtail x$ and hence y is an intermediate output function. From Lemma 1 we know that the number of permutations associated to each such intermediate output function is $(2^n - s)!$ where s is the number of equivalence classes of $\sim$. It remains to estimate the number of intermediate output functions inducing the collision relation $\sim$. By using the well known result regarding the number of solutions of the system of linear equations, we know that the number of vectors satisfying N1 is exactly $2^{n(s-a)}$. If we restrict the solutions such that $y(i) \neq y(j)$ whenever $i \not\sim j$ then the number of solutions is at most $\mathbf{P}(2^n, s - a)$. So $\#\mathbb{P}_n[\sim] \leq\; (2^n - s)! \times \mathbf{P}(2^n, s - a)$. To obtain the lower bound, note that any vector $v \in \mathcal{V}^{**}_{neq}$ is linearly independent with $\mathcal{V}_{eq}$ and hence the rank of $\mathcal{V}_{eq} \cup \{v\}$ is $(a + 1)$. If we remove all solutions satisfying N1 and $v \cdot \mathbf{y} = \mathbf{0}$ for each $v \in \mathcal{V}^{**}_{neq}$, from the set of solutions of N1 then it gives the set of all solutions satisfying N1 and N2. So we have the lower bound.

Corollary 1. $\mathsf{Pr}[\sim_{\Pi} = \sim] \leq \frac{1}{\mathbf{P}(2^n - s + a, a)}$ *where* $a = \mathsf{acc}(\sim)$.

3.2 Generator and Number of Accidents of a Collision Relation

So far we have defined intermediate input function x^π, output function y^π and a collision relation $\sim^\pi$ associated with any permutation π. Now we see that not all collisions (i.e. $x(i) = x(j), i \neq j$) are unexpected. That means there is a set of collision pairs which imply all other collisions independent of the underlying permutation. Generator is a set representing the minimum such set and the number of such collisions are called accident (which is eventually the rank of $\mathcal{V}_{eq}$, the size of a basis of $\mathcal{V}_{eq}$). There can be more than one basis. We choose a special basis in a particular manner so that it uniquely determines the collision relation.

Definition 3. *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 (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* $[1, t]^{(2)} := \{(i, j) : i > j\}$.[4] *The number of accident is defined as* $a = \mathsf{acc}(\sim) := \mathrm{rank}(\mathcal{V}_{eq})$.

[4] $(i, j) \prec (i', j')$ 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

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 $(\mathbf{A}_i^{\sim} - \mathbf{A}_j^{\sim})$ is not linearly independent with $\{\mathbf{A}_{i_c}^{\sim} - \mathbf{A}_{j_c}^{\sim} : c < k\}$. So a relation $\sim$ uniquely determines the generator Gen($\sim$). Now we state that the converse is also true, i.e. a generator uniquely determines a relation. The proof is given in [17]. From now onwards, all missing proofs of our paper are given in [17].

Lemma 3. *Any relation $\sim$ satisfying the necessary condition of Lemma 2 is uniquely determined by its generator* Gen($\sim$). *Hence the number of collision relations with a accident is at most* $\binom{t}{2}^a$.

Corollary 2. $Pr[\mathsf{acc}(\sim_{\Pi}) \geq 2 : \Pi \xleftarrow{*} \mathbb{P}_n] \leq \frac{t^2}{2^n}$ *if* $t < 2^{n/2-1}$.

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 (Corollary 2). All other collisions present in the collision relation are implied from these. For example, CBC can have intermediate collisions for two messages if the messages have common prefix. We see the following example where there are are three collisions among which two of these are unexpected and the third collision can be derived from these two.

Example 1. 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]).

4 Secure Affine Domain Extensions

In Definition 1 we have defined affine domain extension. In this section we study a subclass called secure affine domain extension or SADE. The cipher block

$i > j$, i.e. $(i, j) \in [1, t]^{(2)}$. The notion of smaller and larger for pairs are based on the dictionary order.

chaining message authentication code or CBC-MAC [3, 22] 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, 19]. In the following, we define $y(i) = \pi(x(i))$.

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

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

CBC-MAC [3]: The CBC-MAC is CBC applied to the padded message. Let $\ell(M) = b$ and the input function $x(1) = \alpha_1$ and $x(i) = \alpha_i + y(i-1)$, $2 \le i \le b$.

GCBC* [19]: In case of GCBC*, we consider the messages with $b \ge 2$ and it is defined as $(\mathsf{GCBC}^*)^\pi(M) = \pi\big(\alpha_b + c_\delta \cdot \mathsf{CBC}^\pi(\alpha_1, \ldots, \alpha_{b-1})\big)$ 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]: $\mathsf{OMAC}^\pi(M) = \pi\big(\alpha_b + c_\delta' \cdot \pi(\mathbf{0}) + \mathsf{CBC}^\pi(\alpha_1, \ldots, \alpha_{b-1})\big)$ where $\mathsf{CBC}^\pi(\lambda) = \mathbf{0}$, λ is the empty string. Let $\ell(M) = b+1$ and the input function is

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

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 $x(1) = \mathbf{0}$, $x(i) = \alpha_{i-1} + c_i' \cdot y(1), 2 \le i < b$, and $x(b+1) = \alpha_b + c_\delta \cdot y(1) + \sum_{i=2}^{b} y(i)$.

DAG-based PRF [11, 20]: In [11, 20] a domain extension over a message space $\mathcal{M} = \mathbb{F}_{2^n}^\ell$ is proposed for every non-singular labeled DAG $G = ([1, \ell], \mathcal{E}, c)$ 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 [20], 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 $\mathcal{M} = \mathbb{F}_{2^n}^{\ell}$ and leave readers to verify for other cases.

Definition 4. *A DAG G with ℓ nodes $[1, \ell]$ and a color function $c : \mathcal{E} \to \mathbb{F}_{2^n}$ is called* non-singular *[11] if there exists exactly one source node (in-degree is zero), one sink node ℓ (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 $M = (\alpha_1, \ldots, \alpha_\ell) \in \mathbb{F}_{2^n}^{\ell}$, let $\ell = \ell(M) = \ell$ and $\mathsf{DAG}_G(M) = y(\ell)$ where the input and output functions are

$$x(v) = \alpha_v + \sum_{v' \to v} c(v', v) \cdot y(v'),\ y(v) = \pi(x(v)), 1 \leq v \leq \ell.$$

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 4. *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 [22], 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. Now we characterize a class of PRF secure affine domain extension called secure affine domain extensions.

Definition 5 (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 $y^\pi(i) \neq \mathcal{D}^\pi(M')$, $1 \leq i \leq \ell$ where y^π is the intermediate output function associated with π for the message M.*

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:

CBC-MAC on $\{0, 1\}^*$ is not SADE and it has length extension attack due to forced collision. If we modify the definition of OMAC by choosing $c_0 = \mathbf{1}$ then $\mathcal{D}(\mathbf{0}\|\mathbf{0}) = \pi(\mathbf{0}) = y^{\pi,M}(1)$, $\forall \pi$ 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.

4.1 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 a special collision relation called forced relation. An equivalence relation $\sim^*$ is called **forced relation** if $\mathcal{V}_{\rm eq} = \{\mathbf{0}^t\}$ (i.e. $\mathbf{A}_i^{\sim} = \mathbf{A}_j^{\sim}$ if and only if $i \sim j$) and $\mathcal{V}_{neq}^*$ does not contain the zero vector. So a forced relation clearly satisfies the necessary condition of collision relation and $\#\mathbb{P}_n[\sim^*] > 0$ provided $t(t-1) < 2^n$ (see Lemma 2). So for any $i \not\sim^* j$ there exists a permutation π such that $y^\pi(i) \neq y^\pi(j)$. Since $\mathrm{rank}(\mathcal{V}_{eq}) = 0$, there can exist at most one such collision relation (due to Lemma 3). 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 (converse may not be true). This can be shown as $\mathbf{A}_i^{\sim} = \mathbf{A}_j^{\sim}$ for all $i \sim^* j$ where $\sim = \sim^\pi$. 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 $t(t-1) < 2^n$.

Lemma 5 (Equivalence characterization of SADE). *An affine domain extension is SADE if and only if for any tuple of two distinct messages $\mathbf{M} = (M, M')$, the forced collision relation $\sim^*$ is I-isolated (i.e. $i \not\sim^* j$, for all $j \neq i$, $i \in I = \{t_1 := \ell, t_2 := \ell + \ell'\}$).*

Proof of Theorem 2. Lemma 4 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 $M = (\alpha_1, \ldots, \alpha_\ell)$ and $M' = (\alpha'_1, \ldots, \alpha'_{\ell'})$ be two prefix-free messages, i.e. one is not prefix to other. Suppose $s \geq 0$ with $\alpha_1 = \alpha'_1, \ldots, \alpha_s = \alpha'_s, \alpha_{s+1} \neq \alpha'_{s+1}$. Then $s < \min\{\ell, \ell'\}$ and it is called length of common prefix. Now define a collision relation $\sim$ such that $1 \sim \ell + 1, \ldots, s \sim s + \ell$ 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. □

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

Let $(M_1, \ldots, M_q)$ be any fixed (q, t)-messages with $\ell_i = \ell(M_i)$ and $t_i = \sum_{j=1}^{i} \ell_j$. The final and intermediate index sets are $I = \{t_1, \ldots, t_q := t\}$ and $[1, t] \setminus I$

respectively. To any permutation π, we associate the intermediate input and output function $x^\pi : [1,t] \to \mathbb{F}_{2^n}$ and $y^\pi : [1,t] \to \mathbb{F}_{2^n}$ respectively. We also associate a collision relation $\sim^\pi$ characterizing all collisions on $x^\pi(i)$ values. Now we compute probability of collision between an intermediate input and a final input of Π during the computations of $\mathcal{D}^\Pi(M_1), \ldots, \mathcal{D}^\Pi(M_q)$. More precisely,

$$\epsilon(M_1, \ldots, M_q) := \mathsf{Pr}[x^\Pi(i) = x^\Pi(j), \text{ for some } i \in I \text{ and } j \neq i].$$

It is easy to see that $\epsilon(M_1, \ldots, M_q) \leq \sum_{i<i'} \epsilon(M_i, M_{i'})$ (by using the union bound). The probability that $\sim^\Pi$ has rank two or more corresponding to messages $(M_i, M_{i'})$ is less than $\frac{(\ell_i+\ell_{i'})^4}{2^{2n}}$ (see Corollary2). Since $\mathcal{D}$ is SADE, the forced collision relation does not contribute to the probability $\epsilon(M_i, M_{i'})$. So only accident one collision remains to be considered. Let $N(M_i, M_{i'})$ denote the number of accident one collision relations associated with messages $M_i, M_{i'}$ such that the a final index is related with other index (i.e. either $t_i \sim j$ or $t_{i'} \sim j'$ for some $j \neq t_i$ and $j' \neq t_{i'}$). For any such collision relation $\sim$ we know that $\mathsf{Pr}[\sim^\Pi = \sim] \leq \frac{1}{2^n - \ell_i - \ell_{i'} + 1} \leq \frac{1}{2^n - 2\ell}$ (see Corollary 1) where $\ell = \max_i \ell_i$. Hence

$$\epsilon(M_1, \ldots, M_q) \leq \sum_{1 \leq i < i' \leq q} \left(\frac{N(M_i, M_{i'})}{2^n - 2\ell} + \frac{(\ell_i + \ell_{i'})^4}{2^{2n}}\right)$$

$$< \frac{N(M_1, \ldots, M_q) + 2\ell + 8\ell^3 tq/2^n}{2^n}$$

where $N(M_1, \ldots, M_q) := \sum_{1 \leq i < i' \leq q} N(M_i, M_{i'})$. So if $N(M_i, M_{i'}) \leq c(\ell_i + \ell_{i'})$ for some constant c then $N(M_1, \ldots, M_q) \leq ct(q-1)$. Let $N(t,q) = \max N(M_1, \ldots, M_q)$ where the maximum is taken over all (q,t)-messages. We summarize the above discussion in the following lemma.

Lemma 6. *$\epsilon(M_1, \ldots, M_q) := \mathsf{Pr}[x^\pi(i) = x^\pi(j)$, for some $i \in I$ and $j \neq i] \leq \frac{N(t,q)+2\ell+8\ell^3tq/2^n}{2^n}$. Moreover, if $N(M, M') \leq c(\ell + \ell')$ for some constant c and all messages $M \neq M'$ with $\ell = \ell(M)$ and $\ell(M') = \ell'$ then $\epsilon(M_1, \ldots, M_q) \leq \frac{ct(q-1)+2\ell+8\ell^3tq/2^n}{2^n}$.*

Definition 6. *For a fixed block-wise distinct q-tuple $\mathbf{w} = (w_1, \ldots, w_q)$, a permutation π is said to be* $\mathbf{w}$-regular *if*

type-1: $R(y^\pi|_{\overline{I}})$ (the set of all π-outputs on $\overline{I}$, see Section 2) and $W = \{w_1, \ldots, w_q\}$ are disjoint (i.e. all intermediate outputs associated with the permutation π are different from w_i's) 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.

The above Lemma 6 gives probability of type-2 permutations. A random permutation does not satisfy type-1 property if an intermediate output is from the q-set W. Intuitively, the probability that an intermediate output is from W for first time (in terms of index) has probability less than $\frac{q}{2^n - t}$. Since there are t such intermediate outputs we have the following result.

Lemma 7. $\mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[y^{\Pi}(i) = w_j, \quad \textit{for some } i \in \overline{I} \textit{ and } j] \leq \frac{qt}{2^n - t} \leq \frac{qt+t}{2^n}$.

Now we explain why we have defined **w**-regular permutations. Conditioning on the set of all **w**-regular permutations, the probability that q final outputs of the domain extensions for the messages $M_1, \ldots, M_q$ are $w_1, \ldots, w_q$, is at least $\frac{1}{\mathbf{P}(2^n-1,q)}$. So the conditional decorrelation probability is roughly 2^{-nq}. We can visualize this due to the following reason: Given that Π is **w**-regular, all final intermediate inputs are fresh as they are different from all other intermediate inputs of Π. Moreover, w_j's do not appear in non-final intermediate outputs. Hence the Π-outputs of the final intermediate outputs (which are the output of the domain extensions) can be chosen at random so that they are distinct and different from the intermediate outputs, in particular $w_1, \ldots, w_q$.

Lemma 8. $\mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[y^{\Pi}(t_i) = w_i, 1 \leq i \leq q | \Pi \textit{ is } \mathbf{w}\textit{-regular}] \geq \frac{1}{\mathbf{P}(2^n-1,q)}$

Armed with these lemmas and decorrelation theorem (Theorem 1) we can prove our main result of this section.

Theorem 3. *For any SADE $\mathcal{D}$,* $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \leq \frac{3qt+N(t,q)}{2^n}$ *if* $\ell < 2^{n/3-1}$.

Proof. $\mu_{\mathbf{M},\mathbf{w}} := \mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[y^{\Pi}(t_i) = w_i, 1 \leq i \leq q] \geq \frac{\mathsf{Pr}[\Pi \text{ is } \mathbf{w}\text{-regular}]}{\mathbf{P}(2^n-1,q)}$. Note that if $\ell < 2^{n/3-1}$ then $\frac{8\ell^3 tq}{2^n} \leq tq$ and hence $\mathsf{Pr}[\Pi \text{ is } \mathbf{w}\text{-regular}] \geq 1 - \frac{2tq+t+2\ell+N(t,q)}{2^n}$. Clearly, $t + 2\ell < tq$ and so $\mu_{\mathbf{M},\mathbf{w}} \geq \frac{1-\frac{3qt+N(t,q)}{2^n}}{2^{nq}}$. The result follows from the decorrelation theorem. □

Corollary 3. *If* $N(M,M') \leq c \times (\ell(M) + \ell(M'))$ *for all messages* $M \neq M'$ *and some fixed constant c, then* $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \leq \frac{(3+c)tq}{2^n}$ *if* $\ell < 2^{n/3-1}$.

Theorem 4. *For any SADE $\mathcal{D}$, we have* $\mathbf{Adv}^{\mathrm{prf}}_{\mathcal{D}}(q,t,\ell) \leq \frac{t^2}{2^{n-2}}$.

Proof. The result is immediate from Theorem 3 and Lemma 3 if we have the restriction on ℓ as needed in Theorem 3. To prove the unconditional bound we note that $\mu_{\mathbf{M},\mathbf{w}} \geq \frac{\mathsf{Pr}[\Pi \text{ is } \mathbf{w}\text{-regular}]}{\mathbf{P}(2^n-1,q)} \geq \frac{1-(qt+t)/2^n - t(t-1)/2^n}{2^{nq}}$ since $\mathsf{Pr}_{\Pi \xleftarrow{*} \mathbb{P}_n}[x^{\Pi}(t_i) \neq x^{\Pi}(j) \text{ for all } j \neq i] \geq \mathsf{Pr}[\sim^{\Pi} = \sim^*] \geq (1 - t(t-1)/2^n)$ (Lemma 2) where $\sim^*$ is the forced collision relation. The result follows by using decorrelation theorem. □

6 Improved Security Bounds for Members of $\mathcal{C}$

We provide a sketch (the detail can be found in the full version of the paper [17]) of improved security analysis of members of $\mathcal{C}$. We use the following lemmas proved in [2] (Lemma 12 and Lemma 17 of [2]) and Corollary 3 to provide improved PRF bounds.

Lemma 12 of [2]. For the CBC-MAC and any $M \neq M'$, the number of collision relations of accident one associated with messages M and M' such that $\mathsf{CBC\text{-}MAC}(M) = \mathsf{CBC\text{-}MAC}(M')$ is at most $d'(|\ell(M) - \ell(M')|)$ where $d'(m) = \max_{m' \leq m} d(m)$ and $d(m)$ denotes the number of divisors of m.

Lemma 17 of [2]. For any two prefix-free messages $M \neq M'$, $N(M,M') \leq 8(\ell(M) + \ell(M'))$ for CBC-MAC.

Improved Security Bound for CBC:
By applying Lemma 17 of [2] we have $N(t,q) \leq 4t(q-1)$. 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 $i,j \notin \{1,\ell,\ell+1,t := \ell+\ell'\}$ 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).
Case-A : $\delta_M \neq \delta_{M'}$: In this case one can show easily that $\sim$ is I-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 I-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 reduce to CBC case by dropping the last message block from both the messages.

So we can assume that one of the i,j from the set $\{1,\ell,\ell+1,t\}$ and hence $N(M,M') \leq 8(\ell+\ell')$. 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 $\{(i,j)\}$ be the basis for $\sim$ such that $i,j \notin I$. The first element in $\mathbf{A}^{\sim}_t - \mathbf{A}^{\sim}_{i'}$ must be non-zero (either $c_{\delta'} - c_\delta$ or $c_{\delta'} - 1$ or $c_{\delta'}$), whereas the first element of $\mathbf{A}^{\sim}_i - \mathbf{A}^{\sim}_j$ 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 a 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 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 $N(M,N') \leq 6(\ell+\ell')$. Thus, we have PRF-insecurity bound for OMAC as

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

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 $N(M, M') \leq 2(\ell+\ell')$.

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

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

7 Conclusion and Future Work

We provide a 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.

Acknowledgement This work was supported in part by the National Science Foundation, Grant CNS-0937267. Author also wants to thank Ray Perlner, Liting Zhang and anonymous reviewers for their helpful comments.

References

1. M. Bellare and R. Guérin and P. 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. D. 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. O. Goldreich, S. Goldwasser and S. 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, SIAM Journal on Computing archive Volume **17**, Issue 2 (April 1988), pp 373 - 386, 1988.
14. K. Minematsu and T. Matsushima. Improved Security Bounds for PMAC, TMAC, and XCBC. Fast Software Encryption 2007, Lecture Notes in Computer Science, Volume **4593**, pp 434 - 451, 2007.
15. M. Nandi and A. Mandal. Improved Security Analysis of PMAC. Journal of Mathematical Cryptology, July 2008, Volume **2**, No. 2 : pp 149 - 162.
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. M. Nandi. A Unified Method for Improving PRF Bounds for a Class of Blockcipher based MACs. Cryptology eprint archive 2009/014.
18. M. Nandi. Improved security analysis for OMAC as a pseudorandom function. Journal of Mathematical Cryptology. Volume **3**, Issue 2, pp 133 - 148, 2009.
19. M. Nandi. Fast and Secure CBC-Type MAC Algorithms, FSE 2009, Lecture Notes in Computer Science, Volume **5665**, pp 375 - 393.
20. 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.
21. 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.
22. E. Petrank and C. Rackoff. CBC MAC for real-time data sources. Journal of Cryptology, vol. 13, no. 3, pp. 315 - 338, 2000.
23. Krzysztof Pietrzak. A Tight Bound for EMAC. ICALP (2), 2006, pp 168 - 179.
24. P. Sarkar. Pseudo-Random Functions and Parallelizable Modes of Operations of a Block Cipher, Available in http://eprint.iacr.org/2009/217
25. S. Vaudenay. Decorrelation over infinite domains: the encrypted CBC-MAC case. Communications in Information and Systems (CIS), Volume 1, pp. 75 - 85, 2001.
26. S. 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 $\mathbf{M}(\mathbf{w}) = (M_1, \ldots, M_q)$. The probability is computed over the random choice of Π.

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