



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

# Modeling the Ideal Cipher in Linicrypt

Master Thesis

Frederik Semmel

June 29, 2022

Advisors: Fabio Banfi, Ueli Maurer

Institute of Theoretical Computer Science, ETH Zürich

---

## Abstract

Todo

---

# Contents

---

<b>Contents</b>	<b>ii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Preliminaries</b>	<b>2</b>
2.1 Linicrypt . . . . .	2
2.1.1 Definition of a Linicrypt program . . . . .	2
2.1.2 Type of Adversaries . . . . .	3
2.1.3 Algebraic Representation . . . . .	3
2.1.4 Characterizing Collision Resistance in Linicrypt . . . . .	4
<b>3 Linicrypt with Ideal Ciphers</b>	<b>6</b>
3.1 Revisiting Algebraic Representations . . . . .	6
3.2 Revisiting Collision Structures . . . . .	10
3.3 Adapting the Linicrypt model to use block ciphers . . . . .	13
3.3.1 Algebraic representation for ideal cipher Linicrypt . . . . .	14
3.4 Collision Structure . . . . .	14

## Chapter 1

---

# Introduction

---

Todo

---

## Preliminaries

---

### 2.1 Linicrypt

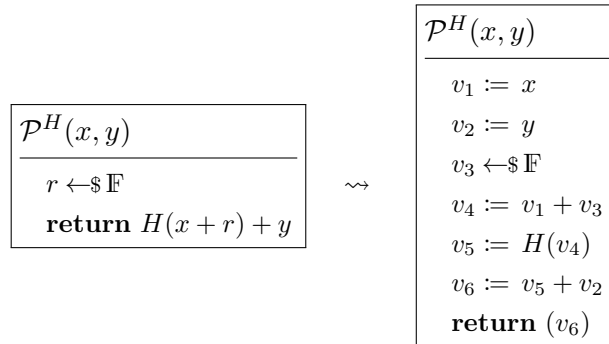
#### 2.1.1 Definition of a Linicrypt program

The Linicrypt model for cryptographic constructions was introduced by Cramer & Rouseluke in [?]. Summarizing the formalization from that paper, a pure Linicrypt program  $\mathcal{P}$  is a straight line program whose intermediate variables are elements in a field  $\mathbb{F}$ . The only allowed operations to create an intermediate variable are:

- Retrieve an input, which is in  $\mathbb{F}$
- Perform a linear combination of existing internal variables
- Call a random oracle  $H : \{0, 1\}^* \times \mathbb{F}^* \rightarrow \mathbb{F}^*$
- Sample from  $\mathbb{F}$  uniformly

The program  $\mathcal{P}$  can output one or more of its variables.

Below is an example of a Linicrypt program  $\mathcal{P}^H$ , written in conventional pseudocode on the left and in explicit Linicrypt on the right.



### 2.1.2 Type of Adversaries

The Linicrypt model only imposes computational restrictions on the constructions, not on the adversaries. Usually one considers arbitrary adversaries  $\mathcal{A}$  that are computationally unbounded, but have bounded access to the random oracle  $H$ . Therefore the behaviour of an adversary is typically described in terms of the number of queries it makes.

### 2.1.3 Algebraic Representation

One of the advantages of restricting the computational model is that one can characterize Linicrypt programs with an algebraic representation. Let  $\mathcal{P}$  be a linicrypt program with intermediate variables  $v_1, \dots, v_n$ . These are sorted in the order in which they are created in the program.

A **base variable** is an intermediate variable which was created by retrieving an input, calling the random oracle  $H$  or sampling from  $\mathbb{F}$ . These are special, because they can be independent of each other. Let **base** be the number of base variables. We denote by  $\mathbf{v}_{\text{base}} \in \mathbb{F}^{\text{base}}$  the column vector consisting of the values that the base variables take in a specific execution of  $\mathcal{P}$ . Here, we order the base variables the same way as the intermediate variables. A **derived variable** is one which is created by performing a linear combination of existing intermediate variables. Note, that derived variables are therefore linear combinations of base variables. As base variables are mostly independent of each other, it makes sense to *model them as linearly independent vectors in  $\mathbb{F}^{\text{base}}$* . The derived variables are then modeled by linear combinations of these vectors.

Let  $v_i$  be an intermediate variable. We define the **associated vector**  $\mathbf{v}_i$  to be the unique row vector such that for every execution of  $\mathcal{P}$  base variables taking the values  $\mathbf{v}_{\text{base}}$ , the variable  $v_i$  has the value  $\mathbf{v}_i \mathbf{v}_{\text{base}}$ . For example, if  $v_i$  is the  $j$ 'th base variable, then  $\mathbf{v}_i = [0 \ \cdots \ 1 \ \cdots \ 0]$ , where the 1 is in the  $j$ 'th position. We follow the convention to write matrices and vectors using a bold font.

The outputs of  $\mathcal{P}$  can be described by a matrix with entries in  $\mathbb{F}$ . Let  $o_1, \dots, o_l \in \{v_1, \dots, v_n\}$  be the output variables of  $\mathcal{P}$ . Then the **output matrix**  $\mathbf{O}$  of  $\mathcal{P}$  is defined by

$$\mathbf{O} = \begin{bmatrix} \mathbf{o}_1 \\ \vdots \\ \mathbf{o}_l \end{bmatrix}.$$

By the definition of the associated vectors  $\mathbf{o}_i$  we have  $\mathbf{O} \mathbf{v}_{\text{base}} = [o_1 \ \cdots \ o_k]^\top$ . The output matrix describes the linear correlations in the output of  $\mathcal{P}$ .

In the same way we also define the **input matrix** of  $\mathcal{P}$ . If  $i_1, \dots, i_k \in \{v_1, \dots, v_n\}$  are

the intermediate variables created by retrieving an input, then we write

$$\mathbf{I} = \begin{bmatrix} \mathbf{i}_1 \\ \vdots \\ \mathbf{i}_k \end{bmatrix}.$$

As  $i_1, \dots, i_k$  are base variables, the rows of  $\mathbf{I}$  are canonical basis vectors. If the Linicrypt program is written such that it first retrieves all its inputs, then  $\mathbf{i}_m$  is simply the  $m$ 'th canonical basis vector.

The input and output matrix describe the linear correlations between input and output of the program with respect to its base variables. But the base variables are not completely independent of each other. In fact, the relationship between the queries and answers to the random oracle  $H$  needs to be captured algebraically. Let  $a_i = H(t_i, (q_1, \dots, q_n))$  be an operation in  $\mathcal{P}$ . The **associated oracle constraint**  $c$  of this operation is the tuple

$$c = \left( t_i, \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix}, \mathbf{a}_i \right) = (t_i, \mathbf{Q}_i, \mathbf{a}_i).$$

This should be interpreted as the requirement that  $\mathbf{a}_i \mathbf{v}_{\text{base}} = H(t_i, \mathbf{Q}_i \mathbf{v}_{\text{base}})$ . We denote the set of all (associated) oracle constraints of  $\mathcal{P}$  by  $\mathcal{C}$ .

As we want the base variables to be linearly independent from each other, we restrict ourselves to Linicrypt programs which don't make multiple calls to the random oracle with the same input. In the language of the algebraic representation: We assume wlog that no two constraints in  $\mathcal{C}$  share the same  $(t, \mathbf{Q})$ .

Wrapping up these definitions, we define the **algebraic representation** of  $\mathcal{P}$  to be the tuple  $(\mathbf{I}, \mathbf{O}, \mathcal{C})$ . A natural question that arises at this point is: Does the algebraic representation determine the behaviour of  $\mathcal{P}$  completely?

The answer is yes, because the algebraic representation does not lose any relevant information about the operations executed in  $\mathcal{P}$ . Informally, this is because the constraints in  $\mathcal{C}$  have such a particular form, which makes it clear in which order the oracles calls have to be executed. Given the order, and using the input matrix, one can determine which variables used in the calls are retrieved from the input and which have to be sampled. Finally the output matrix completely describes how to construct the output from the input, the results of the queries, and randomly sampled values.

The authors of the original paper on Linicrypt [?] establish this result for inputless programs

In chapter 3 we will discuss the

### 2.1.4 Characterizing Collision Resistance in Linicrypt

In a paper by I. McQuoid, T. Swope and M. Rosulek [?, Characterizing Collision and Second-Preimage Resistance in Linicrypt], the authors introduced a characterization

of collision resistance and second-preimage resistance for a class of Linicrypt program based on the algebraic representation.

They identified two reasons why a *deterministic* Linicrypt  $\mathcal{P}$  program can fail to be second-preimage resistant:

1. It is degenerate, meaning that it doesn't use all of its inputs independently
2. It has a collision structure, which means that one can change some intermediate value and compute what the input needs to be to counteract this change

Below are two example Linicrypt programs,  $\mathcal{P}_{\text{deg}}^H$  is degenerate and  $\mathcal{P}_{\text{cs}}^H$  has a collision structure. Note, that you can choose  $w' \neq w$  to be any value, then find a  $x'$  such that the output of  $\mathcal{P}_{\text{cs}}^H$  stays the same, and finally find  $y'$  according to  $w' = x' + y'$ .

$\mathcal{P}_{\text{deg}}^H(x, y)$ <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> $v := x + y$ <b>return</b> $H(v)$	$\mathcal{P}_{\text{cs}}^H(x, y)$ <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> $w := x + y$ <b>return</b> $H(w) + x$
---	--

The precise definition of degenerate and collision structure will be discussed in chapter ??, in a variation that is adapted to the goals of this thesis. The authors show that for any deterministic Linicrypt program which is degenerate or has a collision structure second-preimage resistance (and hence also collision resistance) is completely broken.

The main result of [?] is that they show that the converse of this is also true, for Linicrypt programs which use distinct nonces in each call to the random oracle. That is, if a Linicrypt program is not collision resistant, then it either has a collision structure, or it is degenerate.

Furthermore, checking for degeneracy and existence of a collision structure can be done efficiently.



---

## Linicrypt with Ideal Ciphers

---

### 3.1 Revisiting Algebraic Representations

We have defined the algebraic representation for a Linicrypt program. It consists of the input matrix  $\mathbf{I} \in \mathbb{F}^{k \times \text{base}}$ , the output matrix  $\mathbf{O} \in \mathbb{F}^{l \times \text{base}}$  and the set of constraints  $\mathcal{C} = \{(t_i, \mathbf{Q}_i, \mathbf{a}_i) \mid i = 1, \dots, n\}$ , with  $\mathbf{Q}_i \in \mathbb{F}^{m_i \times \text{base}}$  and  $\mathbf{a}_i \in \mathbb{F}^{1 \times \text{base}}$ . These matrices are such, that if one constraints  $\mathbf{v}_{\text{base}}$  via  $\mathbf{I}\mathbf{v}_{\text{base}} = [i_1 \ \cdots \ i_k]^\top$  for some arbitrary input, the program can compute the rest of the base variables and output  $\mathbf{O}\mathbf{v}_{\text{base}} = [o_1 \ \cdots \ o_l]^\top$ . To simplify the notation we will also write  $(i_1, \dots, i_k)$  for the column vector  $[i_1 \ \cdots \ i_k]^\top$ .

Our goal is to find a better understanding of collision structures. The algebraic representation suggest, that input and output matrix play the same role, they constrain the base variables. Therefore properties of collision resistance should be expressible in terms of the ability to solve the constraints  $\mathcal{C}$  while setting  $\mathbf{O}\mathbf{v}_{\text{base}}$  to some value. This idea will be formalized in this section. A question that arises is: Which combination of matrices with the structure described above correspond in essence to a Linicrypt program? To answer this question, we need to define the terminology more carefully.

**Definition 3.1** (Random oracle constraint). A *random oracle constraint* of dimension  $\text{base}$  taking  $m$  inputs is a tuple  $(t, \mathbf{Q}, \mathbf{a})$  for  $t \in \{0, 1\}^*$ ,  $\mathbf{Q} \in \mathbb{F}^{m \times \text{base}}$  and  $\mathbf{a} \in \mathbb{F}^{1 \times \text{base}}$ .

We call  $t$  the nonce and refer to  $\mathbf{Q}$  and  $\mathbf{a}$  as the query and answer to the random oracle. Usually we just say “a constraint” when the other variables are clear from the context. The constraint  $(t, \mathbf{Q}, \mathbf{a})$  encodes the relationship via the random oracle between the base variables in a program. This semantic meaning of constraints is encoded in the following definition.

**Definition 3.2** (Solution of constraints). Let  $\mathcal{C}$  be a set of constraints of dimension  $\text{base}$ . We say a vector  $\mathbf{v} \in \mathbb{F}^{\text{base}}$  *solves*  $\mathcal{C}$  if  $\mathbf{a}\mathbf{v} = H(t, \mathbf{Q}\mathbf{v})$  for all  $(t, \mathbf{Q}, \mathbf{a}) \in \mathcal{C}$ . Such a  $\mathbf{v}$  is also called a *solution* of  $\mathcal{C}$ .

Clearly, by the definition of the algebraic representation, any instantiation  $\mathbf{v}_{\text{base}}$  of the base variables in a Linicrypt program is a solution to  $\mathcal{C}$ .

Because  $H$  is a well-defined function, and not just any relation, these requirements extend to the constraints.

**Definition 3.3** (Well-defined). *A set of (random oracle) constraints  $\mathcal{C}$  is **well-defined** if for any pair of constraints  $c_i, c_j \in \mathcal{C}$  we have  $(t_i, \mathbf{Q}_i) = (t_j, \mathbf{Q}_j) \implies \mathbf{a}_i = \mathbf{a}_j$ .*

When we use a set of constraints, we will implicitly also require that it is well-defined. Highlighting the function like properties of a constraint  $(t, \mathbf{Q}, \mathbf{a})$ , we will also use the shorthand  $(t, \mathbf{Q}) \mapsto \mathbf{a}$ , or  $\mathbf{Q} \mapsto \mathbf{a}$  if  $t$  is an empty string. This notation was introduced in [?]. If a set of constraints is well-defined, it might still not correspond to a valid Linicrypt program. Consider the set

$$\mathcal{C} = \{[1 \ 0] \mapsto [0 \ 1], [0 \ 1] \mapsto [1 \ 0]\}.$$

Although it is well-defined, it corresponds to the two calls to the random oracle that are not compatible with the randomness assumptions of the oracle. Specifically, such constraints would correspond to a program which does  $y = H(x)$  and  $x = H(y)$ , where  $x, y$  are intermediate variables. This is not possible in the Linicrypt model.

In order to characterize which constraints can be solved by a Linicrypt program, we also need to consider the concept of basis change which was introduced in [?].

**Definition 3.4** (Basis change). *Let  $\mathbf{B}$  be any matrix in  $\mathbb{F}^{\text{base} \times \text{base}}$  and  $\mathcal{P} = (\mathbf{I}, \mathbf{O}, \mathcal{C})$  be a Linicrypt program. We define*

$$\mathcal{CB} := \{(t, \mathbf{QB}, \mathbf{aB}) \mid (t, \mathbf{Q}, \mathbf{a}) \in \mathcal{C}\}.$$

*If  $\mathbf{B}$  is invertible, we say  $(\mathbf{IB}, \mathbf{OB}, \mathcal{CB})$  is the algebraic representation of  $\mathcal{P}$  after applying the **basis change**  $\mathbf{B}$ .*

In the case that  $\mathbf{B}$  is just a permutation of the input variables the transformed representation clearly describes the same program in an informal sense, we have just renamed the input variables. But a more complex basis changes won't change the program either, as is shown below.

Now we can characterize which sets of constraints and input matrices correspond to a Linicrypt program.

**Definition 3.5** (Deterministically Solvable). *Let  $\mathcal{C}$  be a finite set of valid constraints of dimension  $\text{base}$ , and let  $\mathbf{I} \in \mathbb{F}^{k \times \text{base}}$  for some  $k \in \mathbb{N}$ .  $\mathcal{C}$  is **deterministically solvable** fixing  $\mathbf{I}$  if there exists an ordering  $(c_1, \dots, c_n)$  of  $\mathcal{C}$  such that for all  $i = 1, \dots, n$ :*

1.  $\text{rows}(\mathbf{Q}_i) \subset \text{span}(\text{rows}(\mathbf{I}) \cup \{\mathbf{a}_1, \dots, \mathbf{a}_{i-1}\})$
2.  $\mathbf{a}_i \notin \text{span}(\text{rows}(\mathbf{I}) \cup \{\mathbf{a}_1, \dots, \mathbf{a}_{i-1}\} \cup \text{rows}(\mathbf{Q}_i))$

Additionally we require that  $\text{rows}(\mathbf{I}) \cup \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$  form a basis of  $\mathbb{F}^{1 \times \text{base}}$ . We call  $(c_1, \dots, c_n)$  a solution ordering of  $\mathcal{C}$  fixing  $\mathbf{I}$  (or fixing  $\text{rows}(\mathbf{I})$ ).

This is similar to the definition of collision structure in [?]. We will use it to generalize the collision structure and degeneracy attacks from that work.

Intuitively,  $\mathcal{C}$  being deterministically solvable fixing  $\mathbf{I}$  means, that we can constrain  $\mathbf{v}_{\text{base}}$  by  $\mathbf{I}\mathbf{v}_{\text{base}} = (i_1, \dots, i_k)$  for arbitrary inputs  $i_1, \dots, i_k \in \mathbb{F}$ , and then compute each oracle query (condition 1) one by one, without creating a contradiction (condition 2).

If we construct the algebraic representation  $(\mathbf{I}, \mathbf{O}, \mathcal{C})$  of a deterministic Linicrypt program  $\mathcal{P}$ , then  $\mathcal{C}$  is a deterministically solvable set of constraints fixing  $\mathbf{I}$ . Indeed, the solution ordering of  $\mathcal{C}$  fixing  $\mathbf{I}$  can be exactly the order of the corresponding queries in the execution of  $\mathcal{P}$ .

**Lemma 3.6.** *Let  $(\mathbf{I}, \mathbf{O}, \mathcal{C})$  be the algebraic representation of a deterministic Linicrypt program  $\mathcal{P}$  taking  $k$  inputs. Then  $\mathcal{C}$  is deterministically solvable fixing  $\mathbf{I}$ .*

*Proof.* The proof works by induction. Consider a constraint  $(t, \mathbf{Q}, \mathbf{a})$  in  $\mathcal{C}$ . By definition, for some  $i$ ,  $\mathbf{a}$  is the  $i$ 'th canonical basis row vector, and  $\mathbf{Q}$  only has only zeros in the columns to the right of column  $i$ . We can sort  $\mathcal{C} = \{c_1, \dots, c_n\}$ , such that the  $\mathbf{a}$ 's are sorted. Indeed, this is the same order as the associated oracle calls in the execution of  $\mathcal{P}$ .

Clearly condition 2 from definition 3.5 is then fulfilled. Because  $\mathcal{P}$  is deterministic, the queries to the oracle are linear combinations of input variables and results of previous queries. Therefore condition 1 must be fulfilled. Finally, as  $\mathcal{P}$  contains no sampling operation, we know  $\text{base} = k + n$ . The associated vectors  $\text{rows}(\mathbf{I}) = \mathbf{i}_1, \dots, \mathbf{i}_k$  to the input variables are linearly independent, hence it follows that  $\{\mathbf{i}_1, \dots, \mathbf{i}_k, \mathbf{a}_1, \dots, \mathbf{a}_n\}$  is a basis of  $\mathbb{F}^{1 \times \text{base}}$ .  $\square$

The following Lemma shows that the reversed direction also works, meaning that we can recover a Linicrypt program from solvable constraints by applying a basis change. The equivalence up to basis change is not a problem, because any characterization we make about the algebraic representation will always be invariant under basis change.

**Lemma 3.7.** *Let  $\mathcal{C}$  be a set of deterministically solvable constraints fixing  $\mathbf{I} \in \mathbb{F}^{k \times \text{base}}$  for some  $k \in \mathbb{N}$ . Let  $\mathbf{O} \in \mathbb{F}^{\text{out} \times \text{base}}$  be an arbitrary output matrix for some  $\text{out} \in \mathbb{N}$ . Then there is a basis change  $\mathbf{B} \in \mathbb{F}^{\text{base} \times \text{base}}$  and a Linicrypt program  $\mathcal{P}$ , such that  $(\mathbf{I}\mathbf{B}, \mathbf{O}\mathbf{B}, \mathcal{C}\mathbf{B})$  is its algebraic representation.*

*Proof.* Let  $(c_1, \dots, c_n)$  be the solution ordering of  $\mathcal{C}$  fixing  $\mathbf{I} = [\mathbf{i}_1 \ \dots \ \mathbf{i}_k]^\top$ . We choose the new basis for  $\mathbb{F}^{\text{base}}$  as  $(\mathbf{i}_1, \dots, \mathbf{i}_k, \mathbf{a}_1, \dots, \mathbf{a}_n)$ , hence the basis change matrix is defined by

$$\mathbf{B}^{-1} = [\mathbf{i}_1 \ \dots \ \mathbf{i}_k \ \mathbf{a}_1 \ \dots \ \mathbf{a}_n]^\top.$$

In the following we denote by  $\mathbf{e}_i$  the  $i$ 'th canonical row vector. By definition of  $\mathbf{B}$  we have for  $i = 1, \dots, n$

$$\mathbf{I} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k \end{bmatrix} \mathbf{B}^{-1} \quad \text{and} \quad \mathbf{a}_i = \mathbf{e}_{k+i} \mathbf{B}^{-1},$$

which is equivalent to

$$\mathbf{I}\mathbf{B} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k \end{bmatrix} \quad \text{and} \quad \mathbf{a}_i \mathbf{B} = \mathbf{e}_{k+i}.$$

Additionally, as  $\mathcal{C}$  is solvable via the ordering  $(c_1, \dots, c_n)$ , we know that

$$\text{rows}(\mathbf{Q}_i \mathbf{B}) \subset \text{span}(\mathbf{i}_1 \mathbf{B}, \dots, \mathbf{i}_k \mathbf{B}, \mathbf{a}_1 \mathbf{B}, \dots, \mathbf{a}_{i-1} \mathbf{B}) = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{k+i-1})$$

In other words, the query and answer matrices for the constraints in  $\mathcal{CB}$  have the form

$$\begin{aligned} \mathbf{Q}_i \mathbf{B} &= [\lambda_1^i \quad \cdots \quad \lambda_{k+i-1}^i \quad 0 \quad 0 \quad \cdots \quad 0] \quad \text{for } \lambda_j^i \in \mathbb{F} \\ \mathbf{a}_i \mathbf{B} &= [0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0] \end{aligned}$$

Both the input matrix  $\mathbf{I}\mathbf{B}$  and  $\mathcal{CB}$  have the correct form to be the algebraic representation of a program  $\mathcal{P}$  taking  $k$  inputs, and outputting according to  $\mathbf{O}\mathbf{B}$ .  $\square$

In order to abstract the questions about deterministic Linicrypt programs to the level of the algebraic representation, we need the following Lemma.

**Lemma 3.8.** *Let  $\mathcal{P} = (\mathbf{I}, \mathbf{O}, \mathcal{C})$  be a deterministic Linicrypt program. The following two statements are equivalent:*

1. *Running  $\mathcal{P}$  on input  $(i_1, \dots, i_k)$  gives output  $(o_1, \dots, o_l)$ .*

2. *There exists a  $\mathbf{v}$  solving  $\mathcal{C}$  such that  $\mathbf{I}\mathbf{v} = \begin{bmatrix} i_1 \\ \vdots \\ i_k \end{bmatrix}$  and  $\mathbf{O}\mathbf{v} = \begin{bmatrix} o_1 \\ \vdots \\ o_k \end{bmatrix}$ .*

*Proof.* Direction 1.  $\implies$  2. is clear, by definition of the algebraic representation. In fact, such a  $\mathbf{v}$  is unique, as  $\mathcal{P}$  is deterministic.

For the other direction, assume we have such a  $\mathbf{v} \in \mathbb{F}^{\text{base}}$ . If we run  $\mathcal{P}$  on  $(i_1, \dots, i_k)$  we compute the base variables  $\mathbf{v}' \in \mathbb{F}^{\text{base}}$ . By definition of the algebraic representation we have  $\mathbf{I}\mathbf{v}' = (i_1, \dots, i_k) = \mathbf{I}\mathbf{v}$ . Also,  $\mathbf{v}'$  solves  $\mathcal{C}$ , by definition. We will prove that  $\mathbf{v} = \mathbf{v}'$ .

We know  $\mathcal{C}$  is deterministically solvable fixing  $\mathbf{I}$ , hence  $\text{rows}(\mathbf{Q}_1) \in \text{span}\{\text{rows}(\mathbf{I})\}$ . In other words, there is a  $\lambda$  such that  $\mathbf{Q} = \lambda \mathbf{I}$ . Therefore we have  $\mathbf{Q}_1 \mathbf{v}' = \lambda \mathbf{I} \mathbf{v}' = \lambda \mathbf{I} \mathbf{v} = \mathbf{Q}_1 \mathbf{v}$ . Using that both  $\mathbf{v}'$  and  $\mathbf{v}$  solve  $\mathcal{C}$  gives  $\mathbf{a}_1 \mathbf{v}' = H(\mathbf{Q}_1 \mathbf{v}') = H(\mathbf{Q}_1 \mathbf{v}) = \mathbf{a}_1 \mathbf{v}$ . Continuing the induction, we get that  $\mathbf{a}_i \mathbf{v}' = \mathbf{a}_i \mathbf{v}$  for  $i = 1, \dots, n$ . Because  $\text{rows}(\mathbf{I}) \cup \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$  is a basis, the matrix

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

is invertible. Using the assumptions and the results from the induction we have shown that  $\mathbf{B} \mathbf{v}' = \mathbf{B} \mathbf{v}$  and therefore  $\mathbf{v}' = \mathbf{v}$ . Because this means  $\mathbf{O} \mathbf{v}' = \mathbf{O} \mathbf{v} = (o_1, \dots, o_l)$ , we have shown that running  $\mathcal{P}$  on input  $(i_1, \dots, i_k)$  gives output  $(o_1, \dots, o_l)$ .  $\square$

## 3.2 Revisiting Collision Structures

Using this language we can argue about the invertibility and second preimage resistance of a Linicrypt program. The goal of this section is to describe collision structures in a more abstract form, which is easier to adapt to the case of using ideal ciphers instead of random oracles. We will describe collision structures as a program being partially invertible with extra degrees of freedom. As a stepping stone towards this goal, we start by describing a sufficient condition for a Linicrypt program to be invertible.

Let  $\mathcal{P} = (\mathbf{I}, \mathbf{O}, \mathcal{C})$  be a deterministic Linicrypt program taking  $k$  inputs and returning  $l$  outputs. First we note that one can see a deterministic Linicrypt program as a function from its input space to its output space. We associate to  $\mathcal{P}$  the function  $f_{\mathcal{P}} : \mathbb{F}^k \rightarrow \mathbb{F}^l$  which maps  $(i_1, \dots, i_k)$  onto the output of  $\mathcal{P}(i_1, \dots, i_k)$ .

**Lemma 3.9.**  *$\mathcal{P} = (\mathbf{I}, \mathbf{O}, \mathcal{C})$  be a deterministic Linicrypt program. Assume  $\mathcal{C}$  is deterministically solvable fixing  $\mathbf{O}$ . Then  $f_{\mathcal{P}}$  is bijective and  $f_{\mathcal{P}}^{-1}$  is the associated function of a deterministic Linicrypt program which we call  $\mathcal{P}^{-1}$ .*

*Proof.* By Lemma 3.7 we get a Linicrypt program  $\mathcal{P}^{-1}$  with algebraic representation  $(\mathbf{OB}, \mathbf{IB}, \mathcal{CB})$ . Note, that  $\mathbf{O}$  and  $\mathbf{I}$  have the same number of rows, because both  $\text{rows}(\mathbf{I}) \cup \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$  and  $\text{rows}(\mathbf{O}) \cup \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$  are bases of  $\mathbb{F}^{1 \times \text{base}}$ .

By associativity of the matrix product, we have the following equivalence:

$$\begin{aligned} \mathbf{v} \text{ solves } \mathcal{CB} \text{ with } (\mathbf{IB})\mathbf{v} = (i_1, \dots, i_k) \text{ and } (\mathbf{OB})\mathbf{v} = (o_1, \dots, o_l) \\ \Updownarrow \\ \mathbf{B}\mathbf{v} \text{ solves } \mathcal{C} \text{ with } \mathbf{I}(\mathbf{B}\mathbf{v}) = (i_1, \dots, i_k) \text{ and } \mathbf{O}(\mathbf{B}\mathbf{v}) = (o_1, \dots, o_l) \end{aligned}$$

Combining this equivalence with the one from Lemma 3.8 we get: Running  $\mathcal{P}$  on input  $i_1, \dots, i_k$  giving output  $o_1, \dots, o_l$  is equivalent to running  $\mathcal{P}^{-1}$  on input  $o_1, \dots, o_l$  with output  $i_1, \dots, i_k$ . This means that  $f_{\mathcal{P}^{-1}} = f_{\mathcal{P}}^{-1}$ , which is what we needed to prove.  $\square$

For example, the Lemma applies to the following program  $\mathcal{P}$ . The proof constructs the algebraic representation of the inverse program  $\mathcal{P}^{-1}$ .

$\mathcal{P}(x, y)$ <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <b>return</b> $(y + H(x), x)$	$\mathcal{P}^{-1}(w, z)$ <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <b>return</b> $(z, w - H(z))$
---	--

To simplify the following statement about second preimages we introduce a shorthand for all the row vectors contained in a set of constraints  $\mathcal{C}$ . We define

$$\text{rows}(\mathcal{C}) = \bigcup_{(Q, \mathbf{a}) \in \mathcal{C}} \{\text{rows}(\mathbf{Q}) \cup \{\mathbf{a}\}\}.$$

**Lemma 3.10.** *Let  $\mathcal{P} = (\mathbf{I}, \mathbf{O}, \mathcal{C})$  be a deterministic Linicrypt program. Assume we can write  $\mathcal{C} = \mathcal{C}_{\text{fix}} \cup \mathcal{C}_{\text{cs}}$  and  $\mathcal{C}_{\text{cs}}$  is deterministically solvable fixing some  $\mathbf{I}_{\text{cs}}$  such that*

$$\text{span}(\text{rows}(\mathbf{I}_{\text{cs}})) \supset \text{span}(\text{rows}(\mathcal{C}_{\text{fix}}) \cup \text{rows}(\mathbf{O})). \quad (3.1)$$

*Then an attacker can find second preimages with  $\leq n$  queries.*

This Lemma is similar to Lemma 3.9 in the sense that both describe a condition for being able to find a  $\mathbf{v}$  solving  $\mathcal{C}$  while fixing  $\mathbf{O}\mathbf{v} = (o_1, \dots, o_l)$ . The difference is that here we consider the case where we already have a solution  $\mathbf{v}'$  with  $\mathbf{O}\mathbf{v}' = (o_1, \dots, o_l)$ , and require a different one.

Note, that it is crucial that the space on the left is  $\supset$  and not only  $\supseteq$ , as this gives the extra degree of freedom to find a different preimage. It plays the same role as  $\mathbf{q}_{i^*}$  plays in the definition of collision structure from [?]. This characterization includes the case of degeneracy from [?], namely it corresponds to choosing  $\mathcal{C}_{\text{cs}} = \{\}$  and  $\mathbf{I}_{\text{cs}} = \mathbf{1}_{\text{base}}$ , while degeneracy means precisely  $\mathbb{F}^{1 \times \text{base}} \supset \text{span}(\text{rows}(\mathcal{C}) \cup \text{rows}(\mathbf{O}))$ .

*Proof.* Assume we are given a  $\mathbf{v}'$  solving  $\mathcal{C}$  such that  $\mathbf{I}\mathbf{v}' = (i_1, \dots, i_k)$  and  $\mathbf{O}\mathbf{v}' = (o_1, \dots, o_l)$ . The goal is to find a different  $\mathbf{v}$  solving  $\mathcal{C}$  with  $\mathbf{O}\mathbf{v} = (o_1, \dots, o_l)$ . By (3.1) we can constrain  $\mathbf{v}$  to fulfill  $\mathbf{O}\mathbf{v} = \mathbf{O}\mathbf{v}'$  and  $(\mathbf{Q}\mathbf{v} = \mathbf{Q}\mathbf{v}' \text{ and } \mathbf{a}\mathbf{v} = \mathbf{a}\mathbf{v}')$  for every  $(t, \mathbf{Q}, \mathbf{a}) \in \mathcal{C}_{\text{fix}}$ . Even doing so, we still have at least one degree of freedom before we have uniquely determined what  $\mathbf{I}_{\text{cs}}\mathbf{v}$  is. We can therefore choose these further constraints to be arbitrary such that  $\mathbf{v} \neq \mathbf{v}'$ . Because  $\mathcal{C}_{\text{cs}}$  is deterministically solvable fixing  $\mathbf{I}_{\text{cs}}$ , we can uniquely determine all the components of  $\mathbf{v}$  such that they solve  $\mathcal{C}_{\text{cs}}$ . But we chose  $\mathbf{I}_{\text{cs}}\mathbf{v}$  in such a way, that  $\mathbf{v}$  also solves  $\mathcal{C}_{\text{fix}}$ , hence  $\mathbf{v}$  solves  $\mathcal{C} = \mathcal{C}_{\text{cs}} \cup \mathcal{C}_{\text{fix}}$ .

Finally, because  $\mathcal{P}$  is deterministic,  $\mathbf{v} \neq \mathbf{v}'$  implies  $\mathbf{I}\mathbf{v} \neq \mathbf{I}\mathbf{v}'$ , so  $\mathbf{I}\mathbf{v}$  is a second preimage to  $\mathbf{I}\mathbf{v}'$  for the Linicrypt program  $\mathcal{P}$ .  $\square$

The main benefit of this perspective on the collision structure attack is, that it directly describes the collision structure attack as a Linicrypt program. If we apply Lemma

3.7 to  $\mathcal{C}_{cs}$  while fixing  $\mathbf{I}_{cs}$ , the resulting “attack program” is one taking as input the desired output, some values determined by the execution on the given preimage, and some extra values which can be set freely.

The following is the running example from [?].

$\mathcal{P}(x, y, z)$ <hr style="width: 100%;"/> $w = H(x) + H(z) + y$ <b>return</b> $(H(w) + x, H(z))$
---

$$\begin{aligned}
 \mathbf{I} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\
 \mathbf{O} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \\
 \mathbf{q}_1 &= [1 & 0 & 0 & 0 & 0 & 0] \\
 \mathbf{a}_1 &= [0 & 0 & 0 & 1 & 0 & 0] \\
 \mathbf{q}_2 &= [0 & 0 & 1 & 0 & 0 & 0] \\
 \mathbf{a}_2 &= [0 & 0 & 0 & 0 & 1 & 0] \\
 \mathbf{q}_3 &= [0 & 1 & 0 & 1 & 1 & 0] \\
 \mathbf{a}_3 &= [0 & 0 & 0 & 0 & 0 & 1]
 \end{aligned}$$

We set  $\mathcal{C}_{cs} = \{c_1, c_3\}$  and  $\mathcal{C}_{fix} = \{c_2\}$ . Then  $(c_3, c_1)$  is a solution ordering for  $\mathcal{C}_{cs}$  fixing

$$\mathbf{I}_{cs} = [\mathbf{O} \quad \mathbf{q}_2 \quad \mathbf{q}_3]^\top.$$

Note, that for this matrix we have the required condition (3.1), because

$$\text{span}(\text{rows}(\mathbf{I}_{cs})) \supseteq \text{rows}(\mathbf{O}) \cup \text{rows}(\mathcal{C}_{fix}) \cup \{\mathbf{q}_3\},$$

and  $\mathbf{q}_3$  is linearly independent from the rest on the left hand side. Then Lemma 3.7 gives us the transformation  $\mathbf{B}$  such that  $(\mathbf{I}_{cs}\mathbf{B}, \mathbf{IB}, \mathcal{C}_{cs}\mathbf{B})$  is the algebraic representation of some  $\mathcal{P}_{cs}$ . Here we have

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

which leads to the representation:

$$\begin{aligned} \mathcal{C}_{cs}\mathbf{B} &= \left\{ \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \mapsto \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right\} \\ \mathbf{I}_{cs}\mathbf{B} &= \begin{bmatrix} \mathbb{1}_4 & 0 \end{bmatrix} \\ \mathbf{I}\mathbf{B} &= \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

In the form of a Linicrypt program this is:

$\mathcal{P}_{cs}(o_1, o_2, z, q_3)$
$a_3 = H(q_3)$
$v = H(o_1 - a_3)$
<b>return</b> $(o_1 - a_3, q_3 - o_2 - v, z)$

This program computes second preimages to the input  $(x, y, z)$  if we set  $(o_1, o_2) = \mathcal{P}(x, y, z)$  and choose  $q_3$  arbitrarily.

### 3.3 Adapting the Linicrypt model to use block ciphers

In this chapter we modify the Linicrypt model to make use of the ideal cipher model instead of the random oracle model. This means that a Linicrypt program gets access to a block cipher  $\mathcal{E} = (E, D)$  where  $E$  and  $D$  are functions  $\mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$  instead of the hash function  $H : \{0, 1\}^* \times \mathbb{F}^* \rightarrow \mathbb{F}$ . By the definition of a block cipher,  $E_k := E(k, \cdot)$  is a permutation of  $\mathbb{F}$  for all  $k \in \mathbb{F}$  and  $D_k := D(k, \cdot)$  is its inverse. In the ideal cipher model, we assume that the block cipher has no weakness. This is modeled by choosing each permutation  $E_k$  uniformly at random at the beginning of every security game.

The command  $y = E(k, x)$  in a Linicrypt program has to be treated differently from the command  $y = H(k, x)$  when considering collision resistance, because an attacker has access to the deterministic Linicrypt program and both directions of the block cipher  $\mathcal{E} = (E, D)$ . Consider these two programs,  $\mathcal{P}^H$  in standard Linicrypt and  $\mathcal{P}^{\mathcal{E}}$  in ideal cipher Linicrypt.

$\mathcal{P}^H(k, x)$ <hr style="width: 50%; margin: 5px auto;"/> <b>return</b> $H(k, x)$	$\mathcal{P}^{\mathcal{E}}(k, x)$ <hr style="width: 50%; margin: 5px auto;"/> <b>return</b> $E(k, x)$
---	---

While  $\mathcal{P}^H$  is collision resistant, it is trivial to find second preimages for  $\mathcal{P}^{\mathcal{E}}$ . For any  $k' \in \mathbb{F}$  the pair  $(k', D(k', E(k, x)))$  is a second preimage to  $(k, x)$ .

This invertibility property of block ciphers has to be taken into account in both the algebraic representation and the characterization of collision resistance.



### 3.3.1 Algebraic representation for ideal cipher Linicrypt

Let  $\mathcal{P}$  be a ideal cipher Linicrypt program. For each query to  $E$  of the form  $y = E(k, x)$  we define the **associated ideal cipher constraint**  $(E, \mathbf{k}, \mathbf{x}, \mathbf{y})$ . Each query to  $D$  of the form  $x = D(k, y)$ , is associated with the constraint  $(D, \mathbf{k}, \mathbf{y}, \mathbf{x})$ .

As with standard Linicrypt, we want to exclude programs that make unnecessary queries to the block cipher. This way the base variables are linearly independent, except for the dependencies the adversary might introduce by carefully choosing the input. Hence we assume wlog that no two constraints have the same  $(E, \mathbf{k}, \mathbf{x})$  or  $(D, \mathbf{k}, \mathbf{y})$ .

With ideal ciphers there is a second way to make an unnecessary query. That is by first computing  $y = E(k, x)$  and then  $x' = D(k, y)$ . As  $D_k$  is the inverse of  $E_k$  we have  $x = x'$  although  $\mathbf{x}$  and  $\mathbf{x}'$  are linearly independent.

Therefore for all  $\mathbf{k}, \mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}' \in \mathbb{F}^{\text{base}}$  we can assume there are no two constraints  $(E, \mathbf{k}, \mathbf{x}, \mathbf{y})$  and  $(D, \mathbf{k}, \mathbf{y}, \mathbf{x}')$  in  $\mathcal{C}$  for  $\mathbf{x} \neq \mathbf{x}'$ . Neither can there be  $(D, \mathbf{k}, \mathbf{y}, \mathbf{x})$  and  $(E, \mathbf{k}, \mathbf{x}, \mathbf{y}')$  in  $\mathcal{C}$  for  $\mathbf{y} \neq \mathbf{y}'$ .

TODO: Maybe it is simpler with equivalence relation called always colliding queries

$$\begin{aligned} (E, \mathbf{k}, \mathbf{x}, \mathbf{y}) &\sim (E, \mathbf{k}, \mathbf{x}, \mathbf{y}') \\ (E, \mathbf{k}, \mathbf{x}, \mathbf{y}) &\sim (D, \mathbf{k}, \mathbf{y}, \mathbf{x}') \\ (D, \mathbf{k}, \mathbf{y}, \mathbf{x}) &\sim (D, \mathbf{k}, \mathbf{y}, \mathbf{x}') \\ (D, \mathbf{k}, \mathbf{y}, \mathbf{x}) &\sim (E, \mathbf{k}, \mathbf{x}, \mathbf{y}') \end{aligned}$$

And saying that no two constraints in  $\mathcal{C}$  are in the same equivalence class. This might be cleaner, if the equivalence relation used later to analyze repeated nonces case is defined similarly.

## 3.4 Collision Structure

To explain the concept of a collision structure, we will make use of an example. Consider the following Linicrypt program:

$\mathcal{P}_{\text{col},1}^{\mathcal{E}}(a, b, c)$
$w = E(c, b + c) + a$
<b>return</b> $c + E(w, b)$

A second preimage to  $(a, b, c)$  can be found by the following procedure: Choose some  $w' \neq w$ . It will turn out, that even choosing  $w'$  at random, one can calculate what the other base variables need to be such that the output stays the same. As we want  $c + E(w, b) = c' + E(w', b')$  we can again choose any  $b'$  and compute  $c'$ . Finally, we can compute  $a'$  from the equation  $w' = E(c', c' + b') + a'$

One can more easily see that such a procedure is possible by looking at the algebraic representation of  $\mathcal{P}_{\text{col},1}^{\mathcal{E}}$ . In order to highlight which are the base variables, we rewrite the program a bit more explicitly.

$\mathcal{P}_{\text{col},1}^{\mathcal{E}}(a, b, c)$
$k_1 = c$
$x_1 = b + c$
$y_1 = E(k_1, x_1)$
$k_2 = y_1 + a$
$x_2 = b$
$y_2 = E(k_2, x_2)$
<b>return</b> $c + y_2$

The base variables are  $(a, b, c, y_1, y_2)$  and the algebraic representation is

$$\mathbf{O} = [0, 0, 1, 0, 1] \quad \mathcal{C} = \{(E, \mathbf{k}_1, \mathbf{x}_1, \mathbf{y}_1), (E, \mathbf{k}_2, \mathbf{x}_2, \mathbf{y}_2)\}$$

where

$$\begin{aligned} \mathbf{k}_1 &= [0, 0, 1, 0, 0] \\ \mathbf{x}_1 &= [0, 1, 1, 0, 0] \\ \mathbf{y}_1 &= [0, 0, 0, 1, 0] \\ \mathbf{k}_2 &= [1, 0, 0, 1, 0] \\ \mathbf{x}_2 &= [0, 1, 0, 0, 0] \\ \mathbf{y}_2 &= [0, 0, 0, 0, 1]. \end{aligned}$$

We can formulate the previous attack from in terms of the algebraic representation. The task is to find a  $\mathbf{v}'_{\text{base}} = [a', b', c', y'_1, y'_2] \neq \mathbf{v}_{\text{base}}$  such that:

$$\mathcal{P}_{\text{col},1}^{\mathcal{E}}(a', b', c') = \mathbf{O}\mathbf{v}'_{\text{base}} = \mathbf{O}\mathbf{v}_{\text{base}} = \mathcal{P}_{\text{col},1}^{\mathcal{E}}(a, b, c) \quad (3.2)$$

$$y'_1 = \mathbf{y}_1 \mathbf{v}'_{\text{base}} = E(\mathbf{k}_1 \mathbf{v}'_{\text{base}}, \mathbf{x}_1 \mathbf{v}'_{\text{base}}) \quad (3.3)$$

$$y'_2 = \mathbf{y}_2 \mathbf{v}'_{\text{base}} = E(\mathbf{k}_2 \mathbf{v}'_{\text{base}}, \mathbf{x}_2 \mathbf{v}'_{\text{base}}) \quad (3.4)$$

We can fulfill these requirements step by step. First, we constrain  $\mathbf{v}'_{\text{base}}$  by requiring

$$\mathbf{O}\mathbf{v}'_{\text{base}} = \mathbf{O}\mathbf{v}_{\text{base}}.$$

This reduces the dimension of the space of possible solutions for  $\mathbf{v}'_{\text{base}}$  to 4, as  $\mathbf{O}\mathbf{v}_{\text{base}}$  is in the range of  $\mathbf{O}$  by definition. Now note, that neither  $\mathbf{k}_2$  nor  $\mathbf{x}_2$  are in the span of  $\text{rows}(\mathbf{O})$ . Therefore we can choose  $k'_2$  and  $x'_2$  at random such that  $(k'_2, x'_2) \neq (k_2, x_2)$ , and constrain  $\mathbf{v}'_{\text{base}}$  by requiring

$$\mathbf{k}_2 \mathbf{v}'_{\text{base}} = k'_2 \quad \text{and} \quad \mathbf{x}_2 \mathbf{v}'_{\text{base}} = x'_2.$$

Now we can compute  $y'_2 = E(k'_2, x'_2)$  and add constraint (3.4). This constraint is compatible with the previous constraints because  $\mathbf{y}_2$  is not in the span of  $\text{rows}(\mathbf{O}) \cup \{\mathbf{k}_2, \mathbf{x}_2\}$ . The dimension of the subspace of solutions is now 1, as we have added 4 one-dimensional constraints.

Now one only needs to fulfill constraint (3.3). As  $\mathbf{k}_1$  and  $\mathbf{x}_1$  are in the span of  $\text{rows}(\mathbf{O}) \cup \{\mathbf{k}_2, \mathbf{x}_2, \mathbf{y}_2\}$  the intermediate variables  $k'_1$  and  $x'_1$  are uniquely determined. E.g.  $\mathbf{k}_1 = \mathbf{O} - \mathbf{y}_2$  and therefore  $k'_1 = \mathcal{P}_{\text{col},1}^{\mathcal{E}}(a, b, c) - y'_2$ .

Finally, note that  $\mathbf{y}_1 \notin \text{span}(\text{rows}(\mathbf{O}) \cup \{\mathbf{k}_2, \mathbf{x}_2, \mathbf{y}_2\} \cup \{\mathbf{k}_1, \mathbf{x}_1\})$ . Therefore, adding the constraint

$$\mathbf{y}_1 \mathbf{v}'_{\text{base}} = y'_1 = E(k'_1, x'_1)$$

reduces the solution space of possible  $\mathbf{v}'_{\text{base}}$  to a single point in  $\mathbb{F}^{\text{base}}$ . We know that  $\mathbf{v}'_{\text{base}} \neq \mathbf{v}_{\text{base}}$  because  $(k_2, x'_2) \neq (k_2, x_2)$ . The only way to produce different intermediate variables in a deterministic program is to choose different input, hence  $(a', b', c') \neq (a, b, c)$ .

This example would have worked exactly the same if we replaced  $E$  with  $H$ . What follows is an example where the invertibility property of  $E_k$  plays a role.

$\mathcal{P}_{\text{col},2}^{\mathcal{E}}(a, b, c)$	
$k_1 = c$	$\mathbf{O} = [0, 0, 0, 1, 1]$
$x_1 = b$	$\mathbf{k}_1 = [0, 0, 1, 0, 0]$
$y_1 = E(k_1, x_1)$	$\mathbf{x}_1 = [0, 1, 0, 0, 0]$
$k_2 = a$	$\mathbf{y}_1 = [0, 0, 0, 1, 0]$
$x_2 = y_1$	$\mathbf{k}_2 = [1, 0, 0, 0, 0]$
$y_2 = E(k_2, x_2)$	$\mathbf{x}_2 = [0, 0, 0, 1, 0]$
<b>return</b> $y_1 + y_2$	$\mathbf{y}_2 = [0, 0, 0, 0, 1]$

For this program there is a similar procedure to find second preimages. As before, the first constraint is  $\mathbf{O}$ . In this case we can fix  $\mathbf{k}_2 \mathbf{v}'_{\text{base}} = k'_2 = k_2$ ,  $\mathbf{x}_2 \mathbf{v}'_{\text{base}} = x'_2 = x_2$  and  $\mathbf{y}_2 \mathbf{v}'_{\text{base}} = y'_2 = y_2$ . Therefore condition (3.4) is fulfilled trivially. After adding these 4 constraints the solutions space is still 1-dimensional. Note, that  $\mathbf{y}_1 = \mathbf{x}_2$  and it is therefore already fixed at this point, hence to fulfill (3.3) we have to make use of the invertibility property of  $E_k$ . Because

$$\begin{aligned} \mathbf{y}_1 \mathbf{v}'_{\text{base}} &= E(\mathbf{k}_1 \mathbf{v}'_{\text{base}}, \mathbf{x}_1 \mathbf{v}'_{\text{base}}) \\ \iff \mathbf{x}_1 \mathbf{v}'_{\text{base}} &= D(\mathbf{k}_1 \mathbf{v}'_{\text{base}}, \mathbf{y}_1 \mathbf{v}'_{\text{base}}), \end{aligned}$$

we can choose a random  $k'_1 \neq k_1$  and compute  $x'_1 = D(k'_1, y'_1)$  in order to fulfill (3.1). With this fifth constraint we have found a single  $\mathbf{v}'_{\text{base}} \neq \mathbf{v}_{\text{base}}$ .

We want to briefly summarize the conditions for this kind of second preimage attack. Let  $\mathcal{P}'$  denote the execution of

- Some ideal cipher constraints are fulfilled by setting the intermediate variables in  $\mathcal{P}(a', b', c')$  to the same value as in  $\mathcal{P}(a, b, c)$
- There is some constraint for which either the input or the output is undetermined by the previously fixed intermediate variables
- For this constraint and all following constraint one can iteratively call the ideal cipher  $\mathcal{E}$  to set the intermediate variables such that the constraints are fulfilled. This is only possible, if the either the output or the input is undetermined by previously fixed variables.

**Definition 3.11** (Collision structure). Let  $\mathcal{P} = (\mathbf{O}, \mathcal{C})$  be a Linicrypt program with  $|\mathcal{C}| = n$ . A **collision structure** for  $\mathcal{P}$  is an index  $1 \leq i^* \leq n$ , an ordering  $(c_1, \dots, c_n)$  of  $\mathcal{C}$  for  $c_i = (Op_i, \mathbf{k}_i, \mathbf{q}_i, \mathbf{a}_i)$  and a tuple  $(dir_{i^*}, \dots, dir_n)$  for  $dir_i \in \{\text{Forward}, \text{Backward}\}$ , such that the following two conditions hold:

1. Let  $\mathcal{F}_{i^*} = \{\mathbf{k}_1, \dots, \mathbf{k}_{i^*-1}, \mathbf{q}_1, \dots, \mathbf{q}_{i^*-1}, \mathbf{a}_1, \dots, \mathbf{a}_{i^*-1}\}$ . One of the following is true:
  - a)  $dir_{i^*} = \text{Forward}$  and  $\text{span}(\{\mathbf{k}_{i^*}, \mathbf{q}_{i^*}\}) \not\subseteq \text{span}(\mathcal{F}_{i^*} \cup \text{rows}(\mathbf{O}))$
  - b)  $dir_{i^*} = \text{Backward}$  and  $\text{span}(\{\mathbf{k}_{i^*}, \mathbf{a}_{i^*}\}) \not\subseteq \text{span}(\mathcal{F}_{i^*} \cup \text{rows}(\mathbf{O}))$
2. For all  $j \geq i^*$  let  $\mathcal{F}_j = \{\mathbf{k}_1, \dots, \mathbf{k}_{j-1}, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{a}_1, \dots, \mathbf{a}_{j-1}\}$ . One of the following is true:
  - a)  $dir_j = \text{Forward}$  and  $\mathbf{a}_j \notin \text{span}(\mathcal{F}_j \cup \{\mathbf{k}_j, \mathbf{q}_j\} \cup \text{rows}(\mathbf{O}))$
  - b)  $dir_j = \text{Backward}$  and  $\mathbf{q}_j \notin \text{span}(\mathcal{F}_j \cup \{\mathbf{k}_j, \mathbf{a}_j\} \cup \text{rows}(\mathbf{O}))$

TODO: Remove this wordy definition. All the info from here should be integrated into the example

**Definition 3.12** (Collision structure). Let  $\mathcal{P} = (\mathbf{O}, \mathcal{C})$  be a Linicrypt program with  $|\mathcal{C}| = n$ . A **collision structure** for  $\mathcal{P}$  is an index  $1 \leq i^* \leq n$ , an ordering  $(c_1, \dots, c_n)$  of  $\mathcal{C}$  for  $c_i = (Op_i, \mathbf{k}_i, \mathbf{q}_i, \mathbf{a}_i)$  and a tuple  $(dir_{i^*}, \dots, dir_n)$  for  $dir_i \in \{\text{Forward}, \text{Backward}\}$ , such that:

1. The  $i^*$ 'th constraint is unconstrained by the output of  $\mathcal{P}$  and previous fixed constraints. Let  $\mathcal{F} = \{\mathbf{k}_1, \dots, \mathbf{k}_{i^*-1}, \mathbf{q}_1, \dots, \mathbf{q}_{i^*-1}, \mathbf{a}_1, \dots, \mathbf{a}_{i^*-1}\}$  denote the vectors fixed by previous constraints in the ordering.

- a) if  $dir_{i^*} = \text{Forward}$ , the input of the query associated to  $c_{i^*}$  is unconstrained:

$$\text{span}(\{\mathbf{k}_{i^*}, \mathbf{q}_{i^*}\}) \not\subseteq \text{span}(\mathcal{F} \cup \text{rows}(\mathbf{O}))$$

- b) if  $dir_{i^*} = \text{Backward}$ , the output of the query associated to  $c_{i^*}$  is unconstrained:

$$\text{span}(\{\mathbf{k}_{i^*}, \mathbf{a}_{i^*}\}) \not\subseteq \text{span}(\mathcal{F} \cup \text{rows}(\mathbf{O}))$$

2. For all  $j \geq i^*$  the constraint  $c_j$  does not contradict previous constraints. Let  $\mathcal{F} = \{\mathbf{k}_1, \dots, \mathbf{k}_{j-1}, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{a}_1, \dots, \mathbf{a}_{j-1}\}$  denote the vectors fixed by previous constraints in the ordering.

a) if  $\text{dir}_j = \text{Forward}$

$$\mathbf{a}_j \notin \text{span}(\{\mathbf{k}_1, \dots, \mathbf{k}_{j-1}, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{a}_1, \dots, \mathbf{a}_{j-1}\} \cup \{\mathbf{k}_j, \mathbf{q}_j\} \cup \text{rows}(\mathbf{O}))$$

b) if  $\text{dir}_j = \text{Backward}$

$$\mathbf{q}_j \notin \text{span}(\{\mathbf{k}_1, \dots, \mathbf{k}_{j-1}, \mathbf{q}_1, \dots, \mathbf{q}_{j-1}, \mathbf{a}_1, \dots, \mathbf{a}_{j-1}\} \cup \{\mathbf{k}_j, \mathbf{a}_j\} \cup \text{rows}(\mathbf{O}))$$

**Lemma 3.13** (Collision structure gives second preimages). *If collision structure blabla exists for  $\mathcal{P} = (\mathbf{O}, \mathcal{C})$  then blabla with probability 1.*

FindSecondPreimage( $\phantom{}$ )
<hr/> Compute $\mathbf{v}_{\text{base}}$ by executing $\mathcal{P}^{\mathcal{E}}(\mathbf{x})$ ... very similar



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

## Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

---

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

**Title of work** (in block letters):

**Authored by** (in block letters):

*For papers written by groups the names of all authors are required.*

**Name(s):**

**First name(s):**


With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

**Place, date**

**Signature(s)**


*For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.*