

# High Order Side-Channel Security for Elliptic-Curve Implementations

Sonia Belaïd and Matthieu Rivain

CryptoExperts, Paris, France

[sonia.belaid@cryptoexperts.com](mailto:sonia.belaid@cryptoexperts.com)

[matthieu.rivain@cryptoexperts.com](mailto:matthieu.rivain@cryptoexperts.com)

**Abstract.** Elliptic-curve implementations protected with state-of-the-art countermeasures against side-channel attacks might still be vulnerable to advanced attacks that recover secret information from a single leakage trace. The effectiveness of these attacks is boosted by the emergence of deep learning techniques for side-channel analysis which relax the control or knowledge an adversary must have on the target implementation. In this paper, we provide generic countermeasures to withstand these attacks for a wide range of regular elliptic-curve implementations. We first introduce a framework to formally model a *regular algebraic program* which consists of a sequence of algebraic operations indexed by key-dependent values. We then introduce a generic countermeasure to protect these types of programs against advanced single-trace side-channel attacks. Our scheme achieves provable security in the noisy leakage model under a formal assumption on the leakage of randomized variables. To demonstrate the applicability of our solution, we provide concrete examples on several widely deployed scalar multiplication algorithms and report some benchmarks for a protected implementation on a smart card.

**Keywords:** Side-channel countermeasures, elliptic-curve cryptography, masking, noisy leakage model, collision attacks, deep learning-based SCA.

## 1 Introduction

Elliptic Curve Cryptography (ECC) implementations protected with state-of-the-art countermeasures against side-channel attacks might still be vulnerable to advanced attacks that recover significant information on the secret (randomized) scalar with a single leakage trace. Among these attacks, the so-called (collision-based) horizontal attacks are particularly effective [BJPW14, PITM14, PC15, DPN<sup>+</sup>16, NCOS16, BCP<sup>+</sup>19, RIL19, PCBP21]. The principle of these attacks is to defeat the standard randomization techniques without recovering information in the non-randomized variables but by exploiting their collision behavior. This threat is amplified by the recent development of deep learning-based side-channel attacks, which have shown to be especially powerful to generalize and improve advanced side-channel attacks such as template and/or horizontal attacks [CCC<sup>+</sup>19].

In such a context, it is of paramount importance to rely on effective countermeasures achieving quantifiable security. During the last two decades, the cryptographic community has introduced leakage models to build concrete security proofs mainly dedicated to symmetric implementations. Unlike asymmetric implementations for which first-order countermeasures (*i.e.*, sensitive variables are split into two parts, then carefully manipulated) were enough to resist traditional side-channel attacks, symmetric implementations have long been proven to be vulnerable to more complex higher-order attacks [OMHT06, CPRR14] which exploit higher-order statistical moments of the leakage distributions.



Among the leakage models introduced as basis to reason on the security of these symmetric implementations, the most deployed one is undoubtedly the *probing leakage model* introduced by Ishai, Sahai, and Wagner in 2003 [ISW03]. In the  $d$ -probing model, a program is secure if any set of  $d$  intermediate variables is independent from the secrets. While it easily models the exponential complexity of combining an increasing number of noisy variables, the probing model is often criticized by practitioners for being too far from the reality of embedded devices. A more realistic model, referred to as *noisy leakage model* was formalized by Prouff and Rivain in 2013 [PR13]. In this model, every atomic operation in a program is assumed to leak a noisy function  $f(x)$  of its input  $x$ , where the noisy feature is modeled by bounding the statistical distance between the prior and posterior distributions of  $x$  (given the observation  $f(x)$ ). A program is secure in this model if the joint distribution of all the noisy intermediate variables does not reveal more than a negligible information of the secrets.

Now that asymmetric implementations have shown to be also vulnerable to higher-order attacks in the shape of the emerging single-trace attacks (*e.g.*, collision and deep learning based side-channel attacks), a similar reasoning can be applied. The same generic leakage models are expected to be used to reason on the concrete security of these implementations which expect the construction of dedicated advanced formal countermeasures achieving quantifiable security.

**Our Contributions.** This paper presents the first formal model of computation to capture regular implementations of ECC and other exponentiation-based cryptosystems (*e.g.*, RSA, discrete logarithm-based schemes, pairings) in the presence of noisy leakage. In a nutshell, our model of *regular algebraic program* fits any cryptographic algorithm which consists of a sequence of algebraic operations indexed by key-dependent values. We then formally define the noisy leakage model for such programs and introduce an additional assumption on the leakage, the *hiddenness* property, which aims to capture the effect of randomization operations on the leakage of an algebraic variable. Namely, we assume that the leakage of a randomized variable is hard to distinguish from that of a uniform random value. From this ground, we design a generic countermeasure that can be applied to a wide range of regular ECC implementations (and exponentiation-based cryptosystems) and we provide a formal security proof in the noisy leakage model under the hiddenness assumption. Finally, we demonstrate its applicability on several widely used scalar multiplication algorithms and provide concrete benchmarks for a protected implementation on a smart card.

**Organization.** Section 2 gives an overview of the state of affairs regarding side-channel attacks against ECC implementations and the existing countermeasures. Our formal model of computation and leakage is presented in Section 3. We then introduce our generic countermeasure for regular algebraic programs together with its formal security proof in Section 4. Applications and implementation results are finally given in Section 5.

## 2 ECC Implementations and Side-Channel Attacks

Elliptic-curve cryptosystems are a key ingredient of everyday-life applications to provide confidentiality and authentication. They still benefit from exponential security compared to the subexponential security of other systems like RSA. A common feature between the deployed elliptic-curve cryptosystems is their core operation referred to as *scalar multiplication*. It consists in the multiplication of some point  $P$  on an elliptic curve by some (usually secret) scalar  $k$ :

$$kP = \underbrace{P + \dots + P}_{k \text{ times}} .$$

Standard scalar multiplication algorithms are similar to efficient (modular) exponentiation algorithms which scan sequentially the bits of the scalar (or exponent) and apply some operation pattern accordingly. It is well known that concrete implementations of these algorithms are prone to side-channel attacks. We summarize the state of affairs regarding these attacks and the existing protections hereafter.

## 2.1 Regular Scalar Multiplication Algorithms

In a naive binary scalar multiplication algorithm, such as the *double-and-add* algorithm, we process a loop that scans the bits of the scalar and performs a point doubling followed by a conditional point addition. Whenever the current scalar bit equals 1 the addition is performed, while when it equals 0 the addition is skipped. In the same way, a naive binary exponentiation algorithm (the *square-and-multiply*) performs a squaring and a conditional multiplication at each iteration depending on the current exponent bit. The conditional operations performed in these algorithms make possible *simple power analysis* (SPA) attacks exploiting a single leakage trace to recover the secret scalar [KJJ99]. The reason of such a flaw is that point additions and point doublings (or equivalently squarings and multiplications) have different operation flows and hence induce different patterns in a leakage trace. A straight observation of these patterns on a single leakage trace might directly reveal the whole secret scalar.

In order to withstand SPA, scalar multiplication algorithms (or exponentiation algorithms) must be rendered regular, namely they should have a constant control flow, independent of the scalar value. A first solution is to make addition and doubling patterns indistinguishable. This can be achieved by using complete or unified formulae for point addition and point doubling [BJ02, RCB16, BL07a, FJ10] or by the means of side-channel atomicity [CCJ04]. In the latter case, the principle is to build point addition and doubling from the same atomic pattern of field operations (so that a leakage trace is made of a sequence of this single atomic pattern). Another possibility is to render the scalar multiplication algorithm regular by nature, independently of the pattern of field operations in each point operation. Namely, the control flow is identical in each iteration of the scalar multiplication whatever the scalar value. This approach was first followed by Coron in [Cor99] who proposed to perform a dummy addition in the binary algorithm loop whenever the scalar bit equals 0. The obtained *double-and-add-always* algorithm performs a point doubling and a point addition at every loop iteration and the scalar bits are no more distinguishable from a leakage trace.

Extensive literature exists on regular scalar multiplication algorithms, see for instance [Joy07, Joy09, Riv11, FLS14, RCB16]. Examples of well-deployed regular algorithms which enjoy attractive features are the Montgomery ladder [Mon87]<sup>1</sup> and the fixed-window left-to-right scalar multiplication algorithm, respectively displayed in Algorithm 1 and Algorithm 2. In Algorithm 2, we assume that the scalar is encoded in such a way that its digits belong to a basis  $\mathbb{B} \subset \mathbb{Z}$ . We note that in order to avoid timing leakage, one should either use a basis such that  $0 \notin \mathbb{B}$  (see for instance [JT09]) or a complete elliptic-curve addition formula (see for instance [BL07a, RCB16]). Algorithm 1 is one of the most popular regular binary algorithms for scalar multiplication on elliptic curves. It relies on two point registers  $R_0$  and  $R_1$  whose difference is constantly equal to the initial point  $P$  at the end of each iteration. The latter window technique may be preferred when more memory is available. In this context, the scalar is divided into  $w$ -bit elements and each iteration depends on a regular window of  $w$  scalar bits rather than on a single one.

<sup>1</sup>The double-and-add algorithm proposed by Joye [Joy07] is very close to the Montgomery ladder displayed in Algorithm 1. Namely, Steps 5 and 6 are replaced by the instruction  $R_{1-b} \leftarrow 2 \cdot R_{1-b} + R_b$  and the loop is reverted.

**Algorithm 1** Montgomery ladder**Input:**  $P, \mathbf{k} = (k_0, k_1, \dots, k_{n-1})_2$ **Output:**  $Q = [\mathbf{k}]P$ 

1.  $R_0 \leftarrow \mathcal{O}$
2.  $R_1 \leftarrow P$
3. **for**  $i = n - 1$  **downto** 0 **do**
4.    $b \leftarrow k_i$
5.    $R_{1-b} \leftarrow R_{1-b} + R_b$
6.    $R_b \leftarrow 2 \cdot R_b$
7. **end for**
8. **return**  $R_0$

**Algorithm 2** Fixed-window left-to-right scalar multiplication**Input:**  $P, \mathbf{k} = (k_0, k_1, \dots, k_{m-1})_{2^w}$ **Output:**  $Q = [\mathbf{k}]P$ 

1. **forall**  $d \in \mathbb{B}$  **do**  $T[d] = [d]P$
2.  $R \leftarrow T[k_{m-1}]$
3. **for**  $i = m - 1$  **downto** 0 **do**
4.    $R \leftarrow 2^w R$                    //  $w$  doublings
5.    $R \leftarrow R + T[k_i]$
6. **end for**
7. **return**  $R$

We note that whereas such regular algorithms withstand SPA at a first glance, one should also be careful with operations performing data dependent memory accesses. Indeed such algorithms might be vulnerable to cache attacks and other advanced micro-architectural attacks [Pag02, HMM<sup>+</sup>12, BvSY14]. A good practice while implementing such algorithms is to perform operations of the form  $R_{1-b} \leftarrow R_{1-b} + R_b$  in a way that avoids data-dependent memory access. Although (correctly implemented) regular algorithms are secure against SPA (and cache attacks), they might still be vulnerable to side-channel attacks, and in particular to so-called *differential power analysis* (DPA).

## 2.2 Randomization Techniques against DPA

Differential Power Analysis (DPA) belong to the most powerful side-channel attacks [KJJ99]. The principle of these attacks is to target the leakage on intermediate variables of the computation that mix some known but varying data (typically an input plaintext) together with fixed secret data (typically the target secret key). For some early variables in the computation the number of secret bits involved in the expression is limited and can be exhaustively guessed. The different key guesses can then be (in)validated by correlating the predicted variables with the leakage on the actual variables. Different statistical tools can be used to extract such a dependency, see for instance [KJJ99, BCO04, MOP07, BGP<sup>+</sup>11], and scalar multiplication algorithms (in which the secret bits are introduced little by little) are typical targets for these attacks.

Scalar multiplication algorithms might also be vulnerable to so-called *address-bit DPA* [IIT03a] which consists in targeting the direct manipulation of the sensitive scalar bits (or digits) or any address that depends on those bits (typically the memory addresses of the registers  $R_b$  or  $R_{1-b}$  in Algorithm 1).

A sound countermeasure to defeat (address-bit) DPA consists in randomizing the intermediate variables, in order to break the dependency between the key-dependent variables and the leakage. This principle, often referred to as *masking*, has been widely applied to different cryptographic algorithms. Dedicated countermeasures have been exhibited by the cryptographic community to secure scalar multiplication algorithms against both kinds of DPA. We briefly recall hereafter some of the most widely used randomization techniques for ECC implementations, which apply to the scalar, the points (projective) coordinates, and the field elements involved in point operations.

**Randomization of the scalar.** In the context of scalar multiplications on elliptic curves, randomizing the private exponent is the first countermeasure provided by Coron [Cor99] to thwart DPA. In a nutshell, such a scalar multiplication involves a private scalar  $k$  and a public elliptic point  $P$  and is performed over a curve of order  $q$ . Randomizing the private scalar  $k$  consists in adding a random multiple of  $q$  to  $k$ , namely choosing a random number

$r$  and defining:

$$k' \leftarrow k + r \cdot q .$$

Because  $q\mathbf{P}$  equals the point at infinity  $\mathcal{O}$ , the initial multiplication can be performed directly with scalar  $k'$  and yield  $k'\mathbf{P} = (k + r \cdot q)\mathbf{P} = k\mathbf{P} + r\mathcal{O} = k\mathbf{P}$  as output. Variants of this countermeasure consist in dividing the scalar into random shares  $k = k_1 + \dots + k_n \pmod{q}$ , or using the euclidean division  $k = \lfloor k/r \rfloor \cdot r + (k \bmod r)$ .

**Randomization of projective coordinates.** A second countermeasure suggested by Coron to thwart DPA [Cor99] consists in randomizing the projective coordinates of the points. Such coordinates are often used for computation on elliptic curves to save costly inversions. In a nutshell, an elliptic point  $\mathbf{P} = (x, y)$  can be represented through three projective coordinates  $(X : Y : Z)$  such that  $x = X/Z$  and  $y = Y/Z$ . The idea of the randomization here is to multiply each of the three coordinates by a random  $r \neq 0$ , that is defining

$$X' \leftarrow X \cdot r ; \quad Y' \leftarrow Y \cdot r ; \quad Z' \leftarrow Z \cdot r ;$$

(where the above multiplication are performed on the base field). By definition of the projective coordinates, the obtained projective point  $(X' : Y' : Z')$  is still a representative of  $\mathbf{P}$ . This principle further generalizes to the widely used Jacobian coordinates, for which  $(X : Y : Z)$  represents the point  $(x, y)$  with  $x = X/Z^2$  and  $y = Y/Z^3$ , as well as to other inversion-free coordinate systems, *e.g.*, for Edwards curves [HWCD08, BL07b].

**Randomization of field elements.** Finally, randomization techniques might also directly apply to the field elements which compose the point coordinates that are processed in the point operations. In particular, on a prime field elliptic curve, all the intermediate variables of the computation are represented as an integer modulo  $p$ , where  $p$  is the prime characteristic of the base field. A standard randomization technique for such modular integers is to lift them to the ring  $\mathbb{Z}_{hp}$ , for some factor  $h$ , and to randomize them as

$$x' \leftarrow x + r \cdot p \pmod{hp}$$

where  $r$  is some random integer. The modular arithmetic is then performed modulo  $hp$ , thus enabling possible re-randomization from time to time with the addition of a fresh random multiple of  $p$ . At the end of the computation, a reduction modulo  $p$  enables to retrieve the correct result, since the relation  $x' \bmod p = (x + r \cdot p) \bmod p = x$  is maintained throughout the computation.

**Randomization of memory addresses.** Another randomization technique usually applied to protect an implementation against address-bit DPA consists in randomizing the memory addresses (as suggested *e.g.* in [HIM<sup>+</sup>13]) whenever they depend on secret data. Such memory address randomization is especially amenable to a masking of the scalar, as developed in [IIT03b] and later extended in [IISO10]<sup>2</sup>.

## 2.3 Advanced Single-Trace Side-Channel Attacks

Standard DPA and address-bit DPA cannot be directly applied in the presence of a good combination of randomization techniques. However, more advanced side-channel attacks might still be possible. When the scalar is randomized, one cannot use correlation techniques (as in standard DPA) or averaging (as in address-bit DPA) over several executions. In such a context, the adversary has no choice but to rely on *single-trace attacks* which exploit a

<sup>2</sup>Izumi et al. [IISO10] consider a slightly different leakage model in which the attacker can distinguish whether a register is overwritten by the same data.

single leakage trace to infer significant information on the (randomized) scalar. Note that partial information on randomized scalars might still be exploitable, either in the fixed scalar case [SI11, RIL19] or in the nonce scalar case [GRV16], but this partial information must still be significant to retrieve the original scalar (or to break the underlying scheme in the nonce scenario).

Several attack techniques exist which aim at making the most of a single leakage trace to fully recover the (randomized) scalar, or at least a significant part of it. Among these attack techniques, template attacks, (collision-based) horizontal attacks, and deep learning-based attacks are today the main threats of current protected ECC implementations.

The very powerful *template attacks* were first introduced by Chari, Rao and Rohatgi in 2002 [CRR03]. The main idea of such attacks is to estimate the likelihood of a key guess based on a profiling phase where the leakage corresponding to the manipulation of all key guesses were recorded. As all the so-called *profiling attacks*, the adversarial model is quite strong as a copy device is expected to be accessible for the attacker with the ability to choose or at least modify the secret material.

In parallel, *horizontal attacks* [CFG<sup>+</sup>10] exploit algebraic dependencies between several data manipulated at different points in time during a single execution<sup>3</sup>. The detection of such dependencies may strongly reduce the number of key hypotheses. A particular sub-case of horizontal attacks is known as *collision attacks* [SWP03] which directly exploit the presence of colliding values during a computation whose number of occurrences is correlated to the secret. In particular, it has been shown that collision-based horizontal attacks could defeat ECC implementations [HMA<sup>+</sup>08], even when protected by a combination of scalar randomization and point coordinate randomization [BJPW14]. In a nutshell, such attacks consist in detecting collisions within intermediate variables of the algorithms in order to infer information on the scalar bits. For example, in Algorithm 2, one may observe a collision between the point  $T[k_i]$  added to  $R$  in Step 5 and the point  $T[d] = [d]\mathbf{P}$  computed at Step 1, for a given  $d \in [0, 2^w)$ , and hence deduce  $k_i = d$ . As extensively discussed in [BJPW14], this simple principle can be applied to a wide range of ECC implementations and it allows the attacker to defeat standard randomization techniques that do not affect the collision behavior of the algorithm intermediate variables. More advanced horizontal attacks based on cluster analysis are able to recover the scalar in the presence of coordinate re-randomization when collision-based horizontal attacks fail [NC17].

A very recent but already large sequence of works suggests to apply *deep learning techniques*, used so far in numerous fields (e.g., document recognition), to side-channel analysis [CDP17, PSB<sup>+</sup>18, KPH<sup>+</sup>19]. Deep learning-based side-channel attacks are often categorized as profiling attacks as they generally are organized in two steps: a profiling phase in which training side-channel traces are learned and an attacking phase in which real side-channel traces are processed to recover a secret. These attacks happen to be very powerful. Indeed, they can capture (collision-based) horizontal attacks with the advantages of the template attacks but without manually isolating interesting points in time. By detecting horizontal and vertical dependencies in power traces, deep learning-based side-channel attacks currently appear as one of the main threats against cryptographic implementations. In particular, it has been demonstrated in [CCC<sup>+</sup>19] that deep learning-based attacks could defeat an implementation of RSA protected by classical (first-order) randomization countermeasures. Weissbart, Picek, and Batina further demonstrated that convolutional neural networks could be used to perform profiling attacks on EdDSA using Curve25519 [WPB19]. Together with Chmielewski, the authors then extended their work to target even protected implementations (e.g., with projective coordinate re-randomization and scalar randomization) [WCPB20]. In the same vein, Mukhtar et al. [MFK<sup>+</sup>20] used convolutional neural networks to successfully attack elliptic-curve scalar multiplications

<sup>3</sup>*Doubling attacks*, introduced by Fouque and Valette in 2003 [FV03] are close variants but detect dependencies on carefully chosen iterations of two related executions.

protected with Coron’s countermeasures. A concrete methodology to evaluate elliptic-curve cryptosystems with and without countermeasures was even introduced in [MPF<sup>+</sup>22]. Perin, Chmielewski, Batina, and Picek recently demonstrated that they can use deep-learning techniques to correct the missing bits of blinded scalars that are partially recovered with clustering-based horizontal attacks [PCBP21] without even knowing their location. The disadvantage of having to collect many balanced traces for training in the context of deep learning-based side-channel attacks can even be circumvented thanks to advanced data augmentation techniques [MBPK21]. Finally, a recent work shows that combining deep learning techniques and template attacks may result in powerful attacks that outperform state-of-the-art deep learning-based side-channel attacks [WPP22].

## 2.4 Related Work

Boolean masking of the scalar was first proposed by Itoh, Izu and Takenaka [IIT03b] as a memory address randomization in order to thwart address-bit DPA. Unlike previous countermeasures in the same vein, *i.e.*, the exponent splitting countermeasure [CJ01] and the randomized window method [YTT03], which required at least twice the initial execution time, Itoh et al’s proposal only suffers a small overhead. In a nutshell, since the security threat comes from the dependency between the addresses of the registers and the secret bit values, the authors followed the idea of first-order Boolean masking to randomize the latter. Namely, they proposed to `xor` a one-time random value to the secret key to perfectly randomize the register addresses. Other countermeasures are then still necessary to further provide resistance against classical (*i.e.* data-bit) DPA.

This first proposal was later extended by Izumi, Ikegami, Sakiyama and Ohta [ISO10] to additionally circumvent DPA with overwrites. While Itoh et al. focused on side-channel attackers observing intermediate variables or addresses separately, embedded devices may actually leak information on two variables that are consecutively stored in the same register at the same time, which is widely known as *transition leakage*. In order to ensure protection against the latter attacks, Izumi et al. modified Itoh et al’s countermeasure [IIT03b] to make sure that the address of the source register was always different from the address of the destination register. Despite this additional constraint, the authors also managed to achieve better performances in terms of data and time complexity.

Considering the issue of transition leakage further, Heyszl, Mangard, Heinz, Stumpf and Sigl [HMH<sup>+</sup>12] propose to randomize register assignment as a first-order countermeasure. In this regard, our countermeasure can be seen as a generalization of their work with a model that captures location-dependent leakage at high orders. Namely, we also consider randomization of register assignment combined with rerandomization of the underlying variables in order to reach higher orders of side-channel security.

In 2018, Tunstall, Papachristodoulou, and Papagiannopoulos [TPP18] proposed a combination of countermeasures to thwart address-bit DPA and more. They reuse the idea of Boolean masking but also protect the exponent shares against classical (data-bit) DPA. Concretely, the authors achieve security against a broader range of first-order side-channel attacks (exploiting the length of the exponent and other intermediate values) as long as the intermediate states of some registers do not leak. From this assumption, the authors demonstrate that an attacker would need a combination of leakage (in the sense of higher-order side-channel attacks) to recover the exponent.

## 2.5 Our Approach

Our work aims to go one step further by designing and proving a generic countermeasure achieving quantifiable security levels for any masking order  $d$  (*i.e.*, sensitive variables are split into  $d + 1$  shares) while considering the leakage of all the intermediate states. Still

relying on Boolean masking to protect asymmetric implementations, it goes beyond the previous sequence of works on several aspects:

- *Masking order*: Our countermeasure is designed and is proven secure for *any* masking order while the countermeasures from [IIT03b], [IISO10], and [TPP18] are limited to thwart first-order side-channel attacks. Hence the security level can be adapted depending on the actual leakage through the masking order which acts as a security parameter. In particular, attacks exploiting several collisions are covered by our security model.
- *Leakage on algebraic variables*: Several countermeasures are limited to masking the exponent and do not consider any leakage on the algebraic variables (e.g., [IIT03b, IISO10, TPP18]). However, these variables arguably generate most of the leakage in an exponentiation or a scalar multiplication through group operations (in particular multiplications).
- *Formal proof*: While many countermeasures only come with light/informal statements (e.g., [TPP18]) or experiments on some specific devices, such methods are not enough for proving higher-order security. In this paper, we provide the first formal security proof under an up-to-date formal leakage model with well defined assumptions.

Finally, our work puts forward a novel approach to prove the security of elliptic-curve / public-key implementations in the presence of leakage, which is currently missing. We believe our formal framework (with the proper definition of regular algebraic program and the introduction of the hiddenness assumption) offers a fertile ground for further research and formal results on leakage-resistant elliptic-curve / public-key implementations.

## 3 Computation and Leakage Models

### 3.1 Mathematical Background and Notations

All along the paper, the interval  $[i, j]$  or  $[i, j)$  with  $i, j \in \mathbb{Z}$  are understood as integer interval, i.e.  $\mathbb{Z} \cap [i, j]$  and  $\mathbb{Z} \cap [i, j)$ , and the notation  $[j]$  simply refers to  $[1, j]$ .

Let  $p_X : x \mapsto \Pr(X = x)$  denote the probability mass function (pmf) of a discrete random variable  $X$ . Let us further denote by  $\vec{p}_X$  the  $|\mathcal{X}|$ -dimensional vector made of the pmf outputs:  $\vec{p}_X = (p_X(x))_{x \in \mathcal{X}}$ . The statistical distance between two random variables  $X$  and  $Y$  of domain  $\mathcal{X}$  is defined as  $\Delta(X; Y) := \|\vec{p}_X - \vec{p}_Y\|$  with respect to some norm  $\|\cdot\|$ . Unless otherwise specified, we will consider the statistical distance based on the  $L_1$  norm weighted by  $\frac{1}{2}$ , as traditionally used in cryptography, which is

$$\Delta_1(X; Y) := \frac{1}{2} \sum_{x \in \mathcal{X}} |p_X(x) - p_Y(x)| . \quad (1)$$

We shall further make use of the notion of *statistical closeness* inherited from the above distance. Namely, two random variables  $X$  and  $Y$  will be said to be  $\varepsilon$ -close for some  $\varepsilon \geq 0$ , denoted  $X \approx_\varepsilon Y$ , if they verify  $\Delta_1(X; Y) \leq \varepsilon$ . We shall denote by  $X \leftarrow \mathcal{A}(in)$  the action of evaluating a (probabilistic) algorithm on input  $in$  and defining the variable  $X$  as the output. We shall further denote  $X \leftarrow \mathcal{X}$  to define  $X$  as a uniform random element from a finite set  $\mathcal{X}$ .

### 3.2 Computation Model

We introduce hereafter a formal model for what we call a *regular algebraic program* (RAP). Such a program is composed of algebraic operations whose operands and results are read



from and written to an *algebraic* memory (*i.e.* a memory storing algebraic variables) where the latter memory is addressed through a second type of variables (with their own memory): the *indexes*. This way, we can capture any cryptographic algorithm which consists in a sequence of algebraic operations indexed by key-dependent values, such as classical regular algorithms for elliptic curve cryptography (or other exponentiation-based cryptosystems).

We consider an algebraic structure  $\mathbb{A}$  equipped with a set of 2-ary operators. For instance,  $\mathbb{A}$  could be a finite field  $\mathbb{F}_p$  or a ring  $\mathbb{Z}_q$ , in both cases equipped with addition and multiplication operators, or  $\mathbb{A}$  could be an additive group, such as the set of points of an elliptic curve, equipped with an addition operator.

A *regular algebraic program* on  $\mathbb{A}$  is a program that takes as input a tuple  $\mathbf{X} \in \mathbb{A}^{\ell_x}$  and a tuple  $\mathbf{k} \in \mathbb{Z}^{\ell_k}$ , and computes as output a tuple  $\mathbf{Y} \in \mathbb{A}^{\ell_y}$  satisfying  $\mathbf{Y} = f_{\mathbf{k}}(\mathbf{X})$  for an algebraic function  $f_{\mathbf{k}}$  that belongs to some class  $\{f_{\mathbf{k}}; \mathbf{k} \in \mathbb{Z}^{\ell_k}\}$ . For this purpose a regular algebraic program operates with two memories, an algebraic memory and an index memory. The algebraic memory, denoted  $(X_1, X_2, \dots, X_m)$ , is composed of  $m$  cells (aka algebraic variables), each one taking a value in  $\mathbb{A}$ . The index memory, denoted  $(ind_1, ind_2, \dots, ind_n)$ , is composed of  $n$  cells (aka index variables), each one taking a value in  $\mathbb{Z}$ . At the beginning of the computation, the input tuples  $\mathbf{X}$  and  $\mathbf{k}$  are written at the beginning of the algebraic and index memories respectively; at the end of the computation, the output tuple  $\mathbf{Y}$  is read from the beginning of the algebraic memory. The program consists of a sequence of instructions which can be of two natures: index instructions and algebraic instructions. An index instruction performs an operation (integer or Boolean operation) on the index variables. Specifically, an index instruction is of the form

$$ind_{i_1} \leftarrow \text{op}(ind_{i_2}, ind_{i_3}) \quad \text{for } i_1, i_2, i_3 \in [n],$$

$$\text{or } ind_{i_1} \leftarrow ind_{i_2} \quad \text{or } ind_{i_1} \leftarrow cst_{\mathbb{Z}} \quad \text{for } i_1, i_2 \in [n],$$

where  $\text{op}$  denotes some operation which is either the addition (or subtraction) on  $\mathbb{Z}$ , the multiplication on  $\mathbb{Z}$ , the bitwise addition, the bitwise multiplication, and where  $cst_{\mathbb{Z}}$  denotes some constant value in  $\mathbb{Z}$ . An algebraic instruction performs an algebraic operation on algebraic variables, possibly indexed by index variables. Specifically, an algebraic instruction is of the form

$$X_{i_1} \leftarrow \text{Op}(X_{i_2}, X_{i_3}) \quad \text{for } i_1, i_2, i_3 \in [m] \cup \{ind_1, ind_2, \dots, ind_n\}$$

where  $\text{Op}$  denotes an operation of  $\mathbb{A}$ , or of the form

$$X_{i_1} \leftarrow X_{i_2} \quad \text{or } X_{i_1} \leftarrow cst_{\mathbb{A}} \quad \text{for } i_1, i_2 \in [m] \cup \{ind_1, ind_2, \dots, ind_n\}$$

where  $cst_{\mathbb{A}}$  denotes some constant from  $\mathbb{A}$ .

In order to capture randomization techniques in this computation model, we will further consider regular algebraic programs augmented with an additional randomization operation:

$$X_{i_1} \leftarrow R(X_{i_2}) \quad \text{for } i_1, i_2 \in [m] \cup \{ind_1, ind_2, \dots, ind_n\}.$$

From an abstract computation level, this instruction has no effect: it simply copies the algebraic variable  $X_{i_2}$  into  $X_{i_1}$ . On the other hand, for a given implementation, the representation of elements of  $\mathbb{A}$  might be randomizable. This is a case for instance when elements of  $\mathbb{F}_p$  are represented by elements of the ring  $\mathbb{Z}_{hp}$  (for some  $h$ ) which can be (re-)randomized as  $x \leftarrow x + R \cdot p \pmod{hp}$  for some random  $R$ . A reduction modulo  $p$  finally enables to retrieve the right representative. Another example is the group of points of an elliptic curve in Jacobian (or other projective) coordinates. Such a point  $(X : Y : Z)$  can be randomized as  $(X : Y : Z) \leftarrow (XR^2 : YR^3 : ZR)$  for some random  $R$ . These are two classical examples but one could consider many other randomization techniques for different algebraic structures.

A regular algebraic program augmented with such a randomization operation is called a *randomized regular algebraic program* hereafter.

*Remark 1.* We note that index variables could equally be defined over some other (finite) ring, like  $\mathbb{Z}_{2^w}$  to capture a  $w$ -bit architectures, but this difference has no impact on our result.

### 3.3 Noisy Leakage of Regular Algebraic Programs

Informally, the  $\delta$ -noisy leakage model states that during the evaluation of a program  $P$ , each atomic operation leaks a noisy function  $f(x)$  of its input  $x$ . The noisiness of the leakage is captured by assuming that the statistical distance between the distribution of  $x$  and the distribution of  $x$  given an observation  $f(x)$  is bounded by  $\delta$ . In other words, an observation  $f(x)$  only implies a bounded bias on the distribution of  $x$ .

In order to get a formal definition of such a noisy function, one must consider an a priori distribution for  $x$ . In the original definition [PR13] and subsequent generalizations [DDF14, PGMP19], this distribution is naturally taken to be the uniform distribution over the definition set of  $x$ . The statistical distance between  $X$  and  $(X \mid f(X) = y)$  is further averaged over the observation  $f(X) = y$ . This gives the following definition:

**Definition 1** (Noisy function [PR13]). Let  $\mathcal{X}$  be a finite set and let  $\delta \in \mathbb{R}$ . A  $\delta$ -noisy function  $f$  on  $\mathcal{X}$  is a function of domain  $\mathcal{X} \times \{0, 1\}^{\ell_R}$  for some  $\ell_R \in \mathbb{N}$  such that

$$\sum_{y \in \text{Im}(f)} \Pr(Y = y) \cdot \Delta(X; (X \mid Y = y)) \leq \delta, \quad (2)$$

where  $X$  is a uniform random variable over  $\mathcal{X}$  and where  $Y = f(X, R)$  for a uniform random variable  $R$  over  $\{0, 1\}^{\ell_R}$ .

Note that the above definition depends on the notion of statistical distance  $\Delta$ . In the original definition [PR13], the authors suggest to use a statistical distance based on the Euclidean norm (or  $L_2$  norm), while [DDF14] argues that taking the  $L_1$  norm is a more natural choice. It was further recently suggested in [PGMP19] to use a statistical distance notion based on the *relative error*, specifically:

$$\Delta_{\text{RE}}((X \mid Y = y); X) = \max_{x \in \mathcal{X}} \left| \frac{\Pr(X = x \mid Y = y)}{\Pr(X = x)} - 1 \right|. \quad (3)$$

We note that the above notion of distance is not symmetric and the order of the arguments matters:  $\Delta_{\text{RE}}((X \mid Y = y); X) \neq \Delta_{\text{RE}}(X; (X \mid Y = y))$ . Noisy functions based on this distance are referred to as *average relative error* (ARE) noisy functions in [PGMP19] since the relative error in (3) is averaged over the distribution of the leakage  $Y$  according to Definition 1. The authors of [PGMP19] argue that using the above notion does not imply a stronger assumption on the leakage in practice, compared to the definitions based on  $L_1$  and  $L_2$  norms, whereas it has some advantages for security proofs based on reductions to the *random probing model*. For this reason, we shall use this notion in our security proof (see Subsection 4.2) although our result simply generalizes to the other notions.

*Remark 2.* Formally, a noisy leakage function has two arguments: an input  $X$  which corresponds to the variable (or tuple of variables) that leaks, and an input  $R$  which corresponds to the *random tape* of the function, *i.e.*, the randomness from which the noise is generated. This random tape input is freshly sampled from  $\{0, 1\}^{\ell_R}$  for each realization of a leaking variable. For the sake of simplicity, we shall skip the random tape argument from the exposition in the rest of the paper. The noisy leakage of a value  $x \in \mathcal{X}$  (or variable) will hence be denoted  $f(x)$ , to mean “ $f(x, R)$  for a fresh random tape  $R \leftarrow \{0, 1\}^{\ell_R}$ ”, and the resulting  $f(x)$  will be considered as a random variable.

**Application to regular algebraic programs.** We formally define the noisy leakage model hereafter in the context of our computation model. Namely, we consider a (randomized) regular algebraic program  $P$  taking as input a pair of tuples  $(\mathbf{X}, \mathbf{k}) \in \mathbb{A}^{\ell_x} \times \mathbb{Z}^{\ell_k}$ . We shall assume that every index instruction  $ind_{i_1} \leftarrow \text{op}(ind_{i_2}, ind_{i_3})$  leaks a noisy function of the input pair of indexes  $(ind_{i_2}, ind_{i_3})$ . On the other hand, an algebraic instruction

$$X_{i_1} \leftarrow \text{Op}(X_{i_2}, X_{i_3})$$

shall leak a noisy function of the input pair of algebraic values  $(X_{i_2}, X_{i_3})$  together with the triplet of indexes  $(i_1, i_2, i_3)$ . Finally, a randomization operation  $X_{i_1} \leftarrow \text{R}(X_{i_2})$  is assumed to leak a noisy function of  $(X_{i_1}, X_{i_2}, i_1, i_2)$ <sup>4</sup>.

*Remark 3.* We could alternatively assume that the algebraic values and the indexes leak independently (through two noisy leakage functions) but our assumption is more general. Indeed a noisy function of the form

$$f(X_{i_2}, X_{i_3}, i_1, i_2, i_3) = (f_1(X_{i_2}, X_{i_3}), f_2(i_1, i_2, i_3))$$

would be a particular case of our leakage model.

*Remark 4.* Note that the aforementioned leakage model does not directly capture physical defaults (*e.g.* glitches or transitions between variables successively stored in the same register) which are strongly dependent on both the implementation (*e.g.* choice of registers and operations) and the underlying devices. Nevertheless, this model and our subsequent proofs can be easily adapted to such specific scenarios by extending the leakage functions to a group of operations that are jointly leaking. For instance, for glitches, the leakage function would typically apply to a group of operations incoming the same synchronization element. For transition leakage (*e.g.* Hamming distance leakage model), the leakage function would apply to successive operations and/or operations that successively write in the same registers. Such hardware-dependent model extensions would typically work in the same way as the robust probing model generalizes the standard probing model [FGP<sup>+</sup>18].

In order to ease the exposition, we introduce a deterministic algorithm which from a (randomized) regular algebraic program  $P$  and an input pair  $(\mathbf{X}, \mathbf{k})$ , produces a *computation trace* of  $P$  on input  $(\mathbf{X}, \mathbf{k})$ :

$$\mathcal{T} \leftarrow \text{ComputTrace}(P, \mathbf{X}, \mathbf{k}) .$$

This computation trace contains the leaking tuples of all the instructions executed by  $P$  on input  $(\mathbf{X}, \mathbf{k})$ . Concretely, `ComputTrace` initializes  $\mathcal{T}$  to the empty list. Then, it sequentially evaluates each instruction in  $P$  and adds its leaking tuple to  $\mathcal{T}$ . Specifically,

- for an index instruction of the form  $ind_{i_1} \leftarrow \text{op}(ind_{i_2}, ind_{i_3})$ , `ComputTrace` adds the pair  $(ind_{i_2}, ind_{i_3})$  to  $\mathcal{T}$ ;
- for an algebraic instruction of the form  $X_{i_1} \leftarrow \text{Op}(X_{i_2}, X_{i_3})$ , `ComputTrace` adds the tuple  $(X_{i_2}, X_{i_3}, i_1, i_2, i_3)$  to  $\mathcal{T}$ ;
- for a randomization operation of the form  $X_{i_1} \leftarrow \text{R}(X_{i_2})$  `ComputTrace` adds the tuple  $(X_{i_1}, X_{i_2}, i_1, i_2)$  to  $\mathcal{T}$ .

Then, we consider a probabilistic algorithm `LeakageSampler` which on input a computation trace  $\mathcal{T}$  and a  $\delta$ -noisy function family  $\mathcal{F} = \{f_i\}_i$  outputs a list of leakage values:

$$\mathcal{L} \leftarrow \text{LeakageSampler}(\mathcal{T}, \mathcal{F}) ,$$

<sup>4</sup>Note that we do not need to give the attacker the (independent) internal random values that are generated for the randomization operation as the knowledge of a noisy leakage function of both the input and the output is actually stronger.

such that  $\mathcal{L} = \{f_i(d_i)\}$  for  $\mathcal{T} = \{d_i\}_i$ . Specifically, for every element  $d_i$  in the computation trace  $\mathcal{T}$ , the `LeakageSampler` algorithm samples a random tape  $r_i$  uniformly at random from  $\{0, 1\}^{\ell_R}$  and adds the evaluation  $f_i(d_i, r_i)$  to the leakage trace  $\mathcal{L}$ .

We can now formally define the noisy leakage of a program.

**Definition 2** (Noisy leakage). Let  $\mathcal{F}$  be a family of  $\delta$ -noisy functions and let  $P$  be a (randomized) regular algebraic program. The noisy leakage of  $P$  on input  $(\mathbf{X}, \mathbf{k})$  is the distribution  $\mathcal{L}(P, (\mathbf{X}, \mathbf{k}))$  obtained by composing the computation-traces and assign-leakage samplers as

$$\mathcal{L}_{\mathcal{F}}(P, (\mathbf{X}, \mathbf{k})) = \text{LeakageSampler}(\text{ComputTrace}(P, \mathbf{X}, \mathbf{k}), \mathcal{F}) .$$

*Remark 5.* A classical noisy function which fairly well fits the reality of embedded devices could be defined as the sum of the input variable's Hamming weight and a random Gaussian noise. But note that as it is defined, the noisy leakage also captures the well known *random probing leakage* [ISW03, DDF14] which states that during the evaluation of a program  $P$ , each intermediate variable leaks its value with some probability (and leaks nothing otherwise). This specific model would require the noisy function to be defined as the identity of its input variable with some probability (computed from its second input) and to an empty leakage otherwise.

*Remark 6.* We stress that for the considered randomization operations, the random bits sampled in an execution  $X_{i_1} \leftarrow R(X_{i_2})$  can be expressed as a deterministic function of the input-output pair  $(X_{i_1}, X_{i_2})$ . Therefore, modelling the leakage as a noisy function of  $(X_{i_1}, X_{i_2})$  is without loss of generality compared to modeling the leakage as a noisy function of the input and the randomness. Our convention is simply more convenient since we do not need to make the randomness explicit.

We can now define the notion of *leakage resilience* for a randomized regular arithmetic program. Let `Enc` be a probabilistic encoding algorithm that maps an index variable  $\mathbf{k} \in \mathbb{Z}^{\ell_k}$  to an encoded input  $\hat{\mathbf{k}} \leftarrow \text{Enc}(\mathbf{k}) \in \mathbb{Z}^{\ell'_k}$ , with  $\ell'_k \geq \ell_k$ . The notion of leakage resilience is defined with respect to a class of leakage functions  $\mathcal{F}$  (noisy leakage functions in our context) and an encoding `Enc` as follows.

**Definition 3** (Leakage resilience). A randomized regular arithmetic program  $\Pi$  is said to be  $\varepsilon$ -*leakage resilient* against a class of leakage functions  $\mathcal{F}$  with respect to encoding `Enc` if there exists a simulator `Sim` (*i.e.*, a probabilistic algorithm) such that for every  $(\mathbf{X}, \mathbf{k})$ :

$$\text{Sim}(P, \mathbf{X}) \approx_{\varepsilon} \mathcal{L}_{\mathcal{F}}(\Pi, (\mathbf{X}, \text{Enc}(\mathbf{k}))) .$$

### 3.4 Leakage of Randomized Variables

We introduce an additional assumption to capture the effect of the randomization operation  $R$  on the leakage of an algebraic variable. In a nutshell, we require that the leakage of a randomized variable  $R(X)$  is hard to distinguish from the leakage of a random value  $R$  uniformly sampled from  $\mathbb{A}$ . This is formalized hereafter by the notion of  $\varepsilon$ -*hiddenness*.

**Simple definition and examples.** In order to simplify the exposition, we first consider a single noisy function  $f$  with a single input variable from  $\mathbb{A}$ , and generalize the notion later.

**Definition 4** (hiddenness, simple version). Let  $f$  be a noisy function of domain  $\mathbb{A}$  and let  $R$  be a randomization operation. The pair  $(f, R)$  is said to be  $\varepsilon$ -*hiding* if for every  $x \in \mathbb{A}$ , the distribution of  $f(R(x))$  is  $\varepsilon$ -close to the distribution of  $f(U)$  where  $U$  is a uniform random variable over  $\mathbb{A}$ .

In the above definition, the notion of randomization operation should be formally understood as a probabilistic algorithm and  $R(x)$  as a random variable. We illustrate the hiddenness notion with the two following examples:

**Example 1.** Let us further consider the randomization operation which maps an element of  $\mathbb{Z}_p$  to a randomized representation in  $\mathbb{Z}_{hp}$ , namely

$$\mathbf{R}(x) = x + R \cdot p \quad \text{with} \quad R \leftarrow [0, h) .$$

We can say that the pair  $(f, \mathbf{R})$  is  $\varepsilon$ -hiding if the statistical distance between the two distributions

$$f(\mathbf{R}(x)) = f(x + R_1 \cdot p) \quad \text{and} \quad f(R_2) \quad \text{with} \quad \begin{cases} R_1 \leftarrow [0, h) \\ R_2 \leftarrow [0, hp) \end{cases}$$

is upper bounded by  $\varepsilon$ . Or equivalently, the best distinguishing algorithm of the two above distributions (without limit of time and memory) has a maximal success probability of  $\frac{1}{2} + \varepsilon$ .

**Example 2.** A second example is the randomization operation with maps a point defined by its Jacobian coordinates on an elliptic curve of base field  $\mathbb{F}_p$  to a randomized representation, namely

$$\mathbf{R}(\mathbf{P}) = \mathbf{R}((X : Y : Z)) = (X\lambda^2 : Y\lambda^3 : Z\lambda) \quad \text{with} \quad \lambda \leftarrow \mathbb{F}_p^* .$$

We consider a leakage function  $f$  applied on points of this elliptic curve. Similarly, we can say that the pair  $(f, \mathbf{R})$  is  $\varepsilon$ -hiding if the statistical distance between the two distributions

$$f(\mathbf{R}(\mathbf{P})) = f((X\lambda^2 : Y\lambda^3 : Z\lambda)) \quad \text{and} \quad f(\mathbf{U})$$

with  $\lambda \leftarrow \mathbb{F}_p^*$  and  $\mathbf{U} \leftarrow (\mathbb{F}_p)^3$ , is upper bounded by  $\varepsilon$ .

We report hereafter some experimental results which give practical evidence about the hiddenness assumption in the Hamming weight with Gaussian noise leakage model (see Subsection 3.5). We notably show that the statistical distance between the two distributions exponentially decreases with the randomness length  $|h|$  or  $|\lambda|$ .

*Remark 7.* We stress that the practical relevance of the hiddenness assumption jointly depends on the leakage function  $f$  and randomization operation  $\mathbf{R}$ . While the above examples can heuristically achieve hiddenness security for Hamming weight-like leakage functions (still depending on the modulus  $p$  and the coordinates of  $\mathbf{P}$  – see Subsection 3.5), other examples might lead to complete hiddenness insecurity (meaning a non-negligible value for  $\varepsilon$ ). Consider for instance a Boolean masking randomization  $\mathbf{R} : x \mapsto (x \oplus R, R)$  with Hamming weight leakage function. It can be checked that the parity of  $\text{Hw}(x \oplus R, R)$  equals the parity of  $\text{Hw}(x)$  which reveals one full bit of information on  $x$  (hence the hiddenness assumption collapses). Those considerations are reminiscent to recent works proposing to use additive masking in prime fields for low-noise applications [DFS16, MMMS22].

**Complete definition.** We now introduce the complete definition of  $\varepsilon$ -hiddenness for a randomized regular algebraic program. In such a program, after randomization, an algebraic variable may be input of several subsequent operations and hence leak several times. Moreover, we should also capture the leakage of variables defined as functions of one or several randomized variables. In a nutshell, we generalize the  $\varepsilon$ -hiddenness definition to the following cases:

- A leakage function  $f$  associated to a multi-input operation, *i.e.* applying to several algebraic variables and index variables. For such a function  $f$ , the  $\varepsilon$ -hiddenness means

$$f(\mathbf{R}(X), \mathbf{W}, \mathbf{i}) \approx_\varepsilon f(R, \mathbf{W}, \mathbf{i}) \quad \text{with} \quad R \leftarrow \mathbb{A}$$

where  $\mathbf{W}$  denotes a tuple of algebraic variables and  $\mathbf{i}$  denotes a tuple of index variables.

- An algebraic variable whose expression involves a previously randomized variable. In such a case, the  $\varepsilon$ -hiddenness of  $f$  generalizes to

$$f(g(R(X), \mathbf{Z}), \mathbf{W}, \mathbf{i}) \approx_{\varepsilon} f(g(R, \mathbf{Z}), \mathbf{W}, \mathbf{i}) \quad \text{with } R \leftarrow \mathbb{A}$$

where  $\mathbf{Y}$  and  $\mathbf{Z}$  denote tuples of algebraic variables,  $\mathbf{i}$  denotes a tuple of index variables, and  $g$  denotes an algebraic function  $\mathbb{A}^{\ell_g} \rightarrow \mathbb{A}$ .

- Several leakage functions  $f_1, \dots, f_t$  associated to several (multi-input) operations involving the same randomized variable (either as direct input or involved in the expression of an input variable). In such a case, the  $\varepsilon$ -hiddenness  $f_1, \dots, f_t$  is defined as:

$$\begin{aligned} & (f_1(g_1(Y, \dots), \dots), \dots, f_t(g_t(Y, \dots), \dots)) \quad \text{with } Y \leftarrow R(X) \\ & \approx_{t\varepsilon} (f_1(g_1(R, \dots), \dots), \dots, f_t(g_t(R, \dots), \dots)) \quad \text{with } R \leftarrow \mathbb{A} . \end{aligned}$$

Formally, we have the following definition:

**Definition 5** (hiddenness). Let  $\mathcal{F}$  be a family of noisy functions and let  $R$  be a randomization operation. Let  $\mathbf{W}_1, \dots, \mathbf{W}_t, \mathbf{Z}_1, \dots, \mathbf{Z}_t$  be tuples of algebraic variables, let  $\mathbf{i}_1, \dots, \mathbf{i}_t$  be tuples of index variables, and let  $g_1, \dots, g_t$  be algebraic functions. Let  $\mathcal{D}(Y)$  be the distribution defined w.r.t.  $f_1, \dots, f_t \in \mathcal{F}$  and  $Y \in \mathbb{A}$  as:

$$\begin{aligned} \mathcal{D}(Y) = & ( f_1(g_1(Y, \mathbf{Z}_1), \mathbf{W}_1, \mathbf{i}_1) \\ & f_2(g_2(Y, \mathbf{Z}_2), \mathbf{W}_2, \mathbf{i}_2) \\ & \vdots \\ & f_t(g_t(Y, \mathbf{Z}_t), \mathbf{W}_t, \mathbf{i}_t) ) \end{aligned}$$

The pair  $(\mathcal{F}, R)$  is said to be  $\varepsilon$ -*hiding* if for every  $f_1, \dots, f_t \in \mathcal{F}$  and every  $x \in \mathbb{A}$ , the distribution  $\mathcal{D}(R(x))$  is  $(t\varepsilon)$ -close to the distribution  $\mathcal{D}(R)$  where  $R$  is a uniform random variable over  $\mathbb{A}$ .

Although the above assumption might seem strong, it soundly captures the class of collision-based horizontal attacks whenever setting  $\varepsilon = 0$ . Indeed, these attacks do not defeat the randomization and do not recover nor exploit any information on the unrandomized variables. They only observe whether two variables in the control flow of the algorithm collide or not. This collision behavior is not affected by assuming that the target program achieves  $\varepsilon$ -hiddenness even while setting  $\varepsilon$  to 0.

Moreover, we would like to stress that our assumption is not ideal in the sense that there always exists an  $\varepsilon$  for which a set of leakage functions (together with a randomization method) satisfies the hiddenness definition. In what follows, we provide practical evidence that the  $\varepsilon$  might be exponentially small (w.r.t. the randomness size).

### 3.5 Experimental Results for the Hiddenness Assumption

We present hereafter some experimental results to motivate the hiddenness assumption for classic randomization techniques and leakage models. Specifically, we consider the naive, yet practically relevant, leakage model of Hamming weight with additive Gaussian noise:

$$f(x) = \text{Hw}(x) + N \quad \text{with } N \leftarrow \mathcal{N}(0, \sigma_N) , \quad (4)$$

for some standard deviation parameter  $\sigma_N$ .

On the other hand, we consider two classic randomization operations:

- *Modulus randomization*: The randomization operation which maps an element of  $\mathbb{Z}_p$  to a randomized representation in  $\mathbb{Z}_{hp}$ , namely

$$R(x) = x + R \cdot p \quad \text{with} \quad R \leftarrow [0, h) .$$

- *Point randomization*: The randomization operation which maps a point defined by its Jacobian coordinates on an elliptic curve of base field  $\mathbb{F}_p$  to a randomized representation, namely

$$R(\mathbf{P}) = R((X : Y : Z)) = (X\lambda^2 : Y\lambda^3 : Z\lambda) \quad \text{with} \quad \lambda \leftarrow \mathbb{F}_p^* .$$

**Hiddenness for modulus randomization.** For the modulus randomization, we can say that the pair  $(f, R)$  is  $\varepsilon$ -hiding if the statistical distance between the two distributions

$$\mathcal{D}_1 := f(R(x)) = f(x + R_1 \cdot p) \quad \text{and} \quad \mathcal{D}_2 := f(R_2) \quad \text{with} \quad \begin{cases} R_1 \leftarrow [0, h) \\ R_2 \leftarrow [0, hp) \end{cases}$$

is upper bounded by  $\varepsilon$ .

**Hiddenness for point randomization.** For the point randomization, we consider that the leakage function  $f$  is applied to each coordinate of the point. Similarly, we can say that the pair  $(f, R)$  is  $\varepsilon$ -hiding if the statistical distance between the two distributions

$$\mathcal{D}_1 := f(R(\mathbf{P})) = f((X\lambda^2 : Y\lambda^3 : Z\lambda)) \quad \text{and} \quad \mathcal{D}_2 := f(\mathbf{U})$$

with  $\lambda \leftarrow \mathbb{F}_p^*$  and  $\mathbf{U} \leftarrow (\mathbb{F}_p)^3$ , is upper bounded by  $\varepsilon$ .

**Gaussian approximation and Kullback-Leibler divergence.** In order to illustrate the hiddenness assumption on those two specific cases, we estimate the distance between the distributions  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . Since the Hamming weight of a random variable can be fairly approximated by a Gaussian distribution (in particular in the presence of an additive Gaussian noise), we estimate the means  $\mu_1, \mu_2$  and standard deviations  $\sigma_1, \sigma_2$  of the two distributions and approximate  $\mathcal{D}_i \sim \mathcal{N}(\mu_i, \sigma_i)$  for  $i \in \{1, 2\}$ . For the point randomization, the distributions are three-dimensional; we hence get multivariate Gaussian distributions  $\mathcal{D}_i \sim \mathcal{N}(\mu_i, \Sigma_i)$  with mean vectors  $\mu_i \in \mathbb{R}^3$  and covariance matrices  $\Sigma_i \in \mathbb{R}^{3 \times 3}$ .

For each estimated set of parameters for  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , we compute the Kullback-Leibler divergence  $D_{\text{KL}}(\mathcal{D}_1 \parallel \mathcal{D}_2)$  between  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . The KL divergence (a.k.a. relative entropy) is a classic measure of distance between two distributions. For continuous distributions, it is defined as

$$D_{\text{KL}}(\mathcal{D}_1 \parallel \mathcal{D}_2) := \int_{-\infty}^{\infty} p_1(x) \log \left( \frac{p_1(x)}{p_2(x)} \right) dx$$

where  $p_1$  and  $p_2$  respectively denote the pdfs of  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . We chose the KL divergence for convenience. Indeed, the KL divergence between two Gaussian distributions  $\mathcal{D}_i \sim \mathcal{N}(\mu_i, \Sigma_i)$  can be simply evaluated as:

$$D_{\text{KL}}(\mathcal{D}_1 \parallel \mathcal{D}_2) = \frac{1}{2} \left( \log \frac{|\Sigma_2|}{|\Sigma_1|} - d + \text{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)^\top \Sigma_2^{-1} (\mu_2 - \mu_1) \right) \quad (5)$$

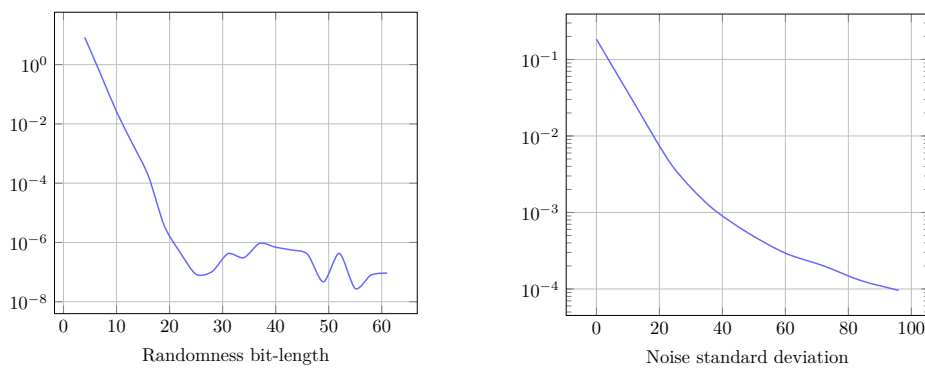
where  $d$  denotes the dimension ( $d \in \{1, 3\}$  in our context) and  $\text{tr}(\cdot)$  the trace function. Moreover,  $D_{\text{KL}}(\mathcal{D}_1 \parallel \mathcal{D}_2)$  gives an upper bound of the statistical distance between  $\mathcal{D}_1$  and  $\mathcal{D}_2$  [CT06]:

$$\Delta_1(\mathcal{D}_1; \mathcal{D}_2) \leq ((\ln 2/2) \cdot D_{\text{KL}}(\mathcal{D}_1 \parallel \mathcal{D}_2))^{\frac{1}{2}} . \quad (6)$$

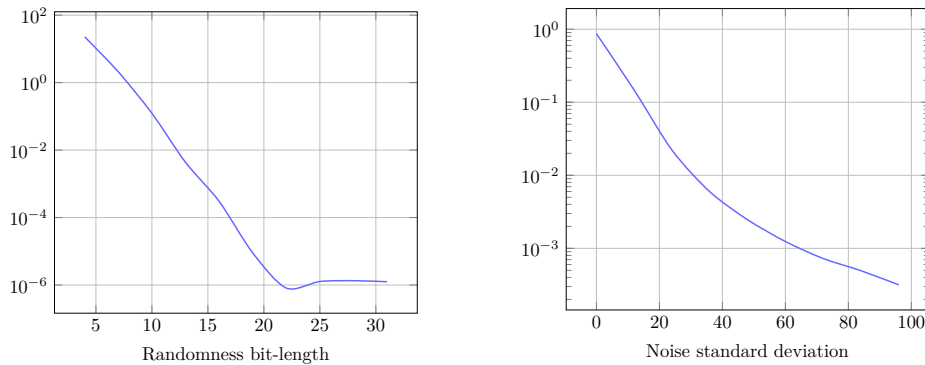
**Experimental setup.** For both studied randomization techniques, we have estimated the distribution parameters to derive an approximation of the KL divergence with respect to

- an increasing randomness bit-length,  $|h| \in [4, 64]$  for modulus randomization,  $|\lambda| \in [4, 32]$  for point randomization, without noise (*i.e.*  $\sigma_N = 0$ ),
- an increasing noise  $\sigma_N \in [0, 100]$  for a fixed (small) randomness bit-length  $|h| = |\lambda| = 8$ .

Our results are depicted in Figure 1 for the modulus randomization and in Figure 2 for the point randomization. The Gaussian means and standard deviations (resp. covariance matrices) and corresponding KL divergences were estimated based on  $10^7$  random samples for each tested set of parameters.



**Figure 1:** Modulus randomization. Left: KL divergence wrt randomness bit-length (no noise). Right: KL divergence wrt noise standard deviation (8-bit randomness).



**Figure 2:** Point randomization. Left: KL divergence wrt randomness bit-length (no noise). Right: KL divergence wrt noise standard deviation (8-bit randomness).

**Observations.** From the left figures, we observe an exponential decrease of the KL divergence – and hence of the statistical distance – with respect to the randomness bit-length. In both cases, the decrease stops beyond the threshold  $D_{\text{KL}}(\mathcal{D}_1 \parallel \mathcal{D}_2) \leq 10^{-6}$ , which is due to the estimation error. We validated this intuition by observing higher thresholds while experimenting with less samples for our estimations. We stress that our experiments using  $10^7$  samples in the different configurations took several days using SageMath on a 3,7 GHz Intel i5 CPU. This exponential decrease with respect to the



randomness length suggests that for sufficiently large randomness, and for the considered leakage function  $f$  (Hamming weight model), the two distributions are indistinguishable, *i.e.* the hiddenness assumption holds for a negligible  $\varepsilon$ .

From the right figures, we further observe that the noise standard deviation seems to have a sub-exponential impact on the KL divergence. This is the expected behavior since it can be checked that the contribution of the additive Gaussian noise  $N$  has a  $\Theta(1/\sigma_N^2)$  impact to the formula (5) – and hence a  $\Theta(1/\sigma_N)$  impact to the statistical distance.

**Potential bias vs. modulus and point coordinates.** For the modulus randomization, we used the ANSSI elliptic curve prime modulus

$$p_{\text{ANSSI}} = \text{0xF1FD178C0B3AD58F10126DE8CE42435B3961ADBCABC8CA6DE8FCF353D86E9C03} \quad (7)$$

We deliberately avoided using a sparse prime such as the NIST P-256 prime

$$p_{\text{NIST}} = 2^{256} - 2^{224} + 2^{192} + 2^{96} - 1 .$$

Indeed, for such a prime, the effect of the randomness does not propagate to all the bits whenever  $|h| < 96$  which implies a direct bias in the distribution  $\mathcal{D}_1 = f(x + R_1 \cdot p)$  compared to  $\mathcal{D}_2 = f(R_2)$ . To avoid this bias, we have to use a non-sparse prime  $p$  or a randomness bit-length which is greater than the longest chain of 0's or 1's in the prime binary expansion (*i.e.* greater than 96 for  $p_{\text{NIST}}$ ). For the distribution  $\mathcal{D}_1 = f(x + R_1 \cdot p)$  we arbitrarily chose  $x = 0$ , but we observed similar results for other values of  $x$ .

For the point randomization, we used the NIST elliptic curve base point [NIS13], defined by the coordinates:

$$\begin{aligned} x_{\mathbf{P}} &= \text{0x6B17D1F2E12C4247F8BCE6E563A440F277037D812DEB33A0F4A13945D898C296} \\ y_{\mathbf{P}} &= \text{0x4FE342E2FE1A7F9B8EE7EB4A7C0F9E162BCE33576B315ECECBB6406837BF51F5} \end{aligned}$$

Since the randomization works on Jacobian coordinates  $(X : Y : Z)$  we further had to define a  $Z$ -coordinate. To avoid a direct bias in the distribution  $\mathcal{D}_1 = f((X\lambda^2 : Y\lambda^3 : Z\lambda))$  compared to  $\mathcal{D}_2 = f(\mathbf{U})$ , the Jacobian coordinates should not be sparse or particularly small compared to  $p$ . The obvious choice of  $(X, Y, Z) = (x_{\mathbf{P}}, y_{\mathbf{P}}, 1)$  implies such a direct (and strong) bias. For this reason, we picked a random  $Z$  in our experiments. Specifically  $Z$  is randomly drawn from  $\mathbb{F}_p$  at the beginning of the experiment and we then fix the coordinates  $(X, Y, Z) = (x_{\mathbf{P}} \cdot Z^2, y_{\mathbf{P}} \cdot Z^3, Z)$  for all the generated samples.

We stress that for each randomization technique, the direct bias exhibited above can be avoided by taking a large enough randomness – typically  $|h| = |p|$  for the modulus randomization and  $|\lambda| = |p|$  for point randomization – but as illustrated in our experiments, shorter randomness seems to be sufficient in the absence of such pathological cases.

The above experiments provide evidence of the practical relevance of the hiddenness assumption. We leave the assessment of this assumption (and the underlying  $\varepsilon$ ) in various contexts (depending on the leakage model and on the randomization technique) as an open research question. This question is related to the design of side-channel attacks that exploit the leakage of randomized variables to recover information on the unrandomized variables (beyond the collision behavior). This issue has not been widely investigated, presumably due to the hardness of exploiting such leakage. To the best of our knowledge, only one recent work studies how to defeat the randomization of projective (and Jacobian) coordinates [ADP<sup>+</sup>20]. The authors investigate strategies to recover the random value  $\lambda$  by exploiting the leakage of the randomization operation  $\mathbf{R}(\mathbf{P})$  assuming a known point  $\mathbf{P}$  (typically the initial point randomization). Among many contributions, this work illustrates the difficulty to efficiently exploit the leakage on point randomization even in the favorable context of 8-bit devices.

## 4 A Generic and Provably-Secure Countermeasure

### 4.1 The Generic Countermeasure

We consider a particular form of regular algebraic program which encompasses a large class of algorithms including elliptic-curve scalar multiplications, ring or field exponentiations, or pairing evaluations. This generic algorithm (see [Algorithm 3](#)) takes as input a tuple of algebraic variables  $\mathbf{X}$  and a secret index variable  $\mathbf{k} = (k_0, \dots, k_{n-1})$  for some  $n \in \mathbb{N}$ . We assume that each coordinate of the input secret variable  $\mathbf{k}$  is a  $w$ -bit integer, that is  $k_i \in [0, 2^w)$  for every  $i \in [0, n)$ . The algorithm is composed of an initialization procedure `preprocess` that initializes the algebraic variables  $X_1^0, X_2^0, \dots, X_m^0$  from the input  $\mathbf{X}$ , a main loop that applies a sequence of operations on the algebraic variables, and a final procedure `postprocess` to generate the output. In the main loop, the procedure `sequence` implements a sequence of operations between algebraic variables of constant indexes, and the procedure `indexes` derives a set of indexes from the current index variable  $k_i$  from which the algebraic variables entering the next iteration are selected. In the following, we shall assume that this function is linear with respect to the bitwise addition. Specifically, in Step 4, each  $b_j$  is defined as a linear combination of the  $w$  bits of  $k_i$ .

---

#### Algorithm 3 Generic scalar multiplication (or exponentiation) algorithm

---

**Input:**  $\mathbf{X}, \mathbf{k} = (k_0, k_1, \dots, k_{n-1})$

**Output:**  $\mathbf{Y} = f_{\mathbf{k}}(\mathbf{X})$

1.  $(X_1^0, X_2^0, \dots, X_m^0) \leftarrow \text{preprocess}(\mathbf{X})$
  2. **for**  $i = 0$  **to**  $n - 1$  **do**
  3.  $((X_1^0, X_1^1), (X_2^0, X_2^1), \dots, (X_m^0, X_m^1)) \leftarrow \text{sequence}(X_1, X_2, \dots, X_m)$
  4.  $(b_1, b_2, \dots, b_m) \leftarrow \text{indexes}(k_i)$
  5.  $(X_1^0, X_2^0, \dots, X_m^0) \leftarrow (X_1^{b_1}, X_2^{b_2}, \dots, X_m^{b_m})$
  6. **end for**
  7.  $\mathbf{Y} \leftarrow \text{postprocess}(X_1^0, X_2^0, \dots, X_m^0)$
  8. **return**  $\mathbf{Y}$
- 

For the sake of clarity, we do not strictly stick to the notations introduced in [Section 3](#) for the algebraic and index memories. However, it should be clear from the algorithm description that

- the  $X$  notations are used for algebraic variables whereas the  $k$  and  $b$  notations are used for index variables,
- the procedures `preprocess`, `sequence` and `postprocess` only perform algebraic operations with constant indexes,
- the procedure `indexes` only perform index operations,
- Step 5 performs a sequence of  $m$  copies of algebraic variables indexed by index variables.

In the following, we shall denote  $N_{pre}$ ,  $N_{post}$ ,  $N_{seq}$  the number of algebraic operations performed by the `preprocess`, `postprocess` and `sequence` procedures respectively. We will further consider that the `indexes` procedure is composed of  $m$  index operations. The total number  $N_P$  of operations in [Algorithm 3](#) is hence given by:

$$N_P = N_{pre} + N_{post} + n \cdot (N_{seq} + 2m) . \quad (8)$$

*Remark 8.* In the description of [Algorithm 3](#), we implicitly assume the same `sequence` procedure for all the iterations but we could generalize to different `sequencei` procedures without effect on our result. Couples of variables are used in Step 3 to record both scenarios

for each input, but notice that in practice a single extra register can be enough to treat both cases for each of the  $m$  input variables.

We show in Section 5 that this general algorithm can be applied to the most widely used elliptic-curve scalar multiplication algorithms, such as the Montgomery ladder [Mon87], the Joye ladder [Joy07], the full signed binary algorithm [Riv11], the fixed window double-and-add algorithm [HMOV04].

**Encoding of the key.** Our generic countermeasure is based on a Boolean sharing of  $\mathbf{k}$  of *masking order*  $d \in \mathbb{N}$  (which acts as security parameter of the countermeasure). Specifically,  $\mathbf{k}$  is encoded as

$$\widehat{\mathbf{k}} \leftarrow \text{Enc}(\mathbf{k}) := ((k_0^j)_{0 \leq j \leq d}, (k_1^j)_{0 \leq j \leq d}, \dots, (k_{n-1}^j)_{0 \leq j \leq d}) \quad (9)$$

where the encoding algorithm  $\text{Enc}$  randomly samples

$$k_i^j \leftarrow [0, 2^w) \quad \text{for every } 0 \leq i \leq n-1 \text{ and } 1 \leq j \leq d \quad (10)$$

and define

$$k_i^0 \leftarrow k_i \oplus k_i^1 \oplus \dots \oplus k_i^d \quad \text{for every } 0 \leq i \leq n-1. \quad (11)$$

**Protected algorithm.** Our generic countermeasure against horizontal side-channel attacks is described in Algorithm 4. The function  $\text{R}$  applies the randomization operation formalized in our randomized regular algebraic program model. The function  $\text{swap}$  performs a conditional swapping of the two input variables  $X^0$  and  $X^1$  depending on the third input  $b$ , so that it returns  $(X^b, X^{b \oplus 1})$  (the variables are swapped iff  $b = 1$ ). In the RAP model, this function can be implemented by the following sequence of instructions with the use of two extra algebraic variables  $R_0$  and  $R_1$ :

1.  $(R_0, R_1) \leftarrow (X^0, X^1)$
2.  $X^0 \leftarrow R_b$
3.  $X^1 \leftarrow R_{b \oplus 1}$

In a nutshell, Algorithm 4 performs the same computation as Algorithm 3 but the selection of the variables in Step 5 is protected by a Boolean masking of order  $d$ . By linearity of the indexes procedure, we have

$$b_\ell = b_\ell^0 \oplus b_\ell^1 \oplus \dots \oplus b_\ell^d, \quad (12)$$

which implies that

$$\text{swap}(X_\ell^0, X_\ell^1, b_\ell) = \text{swap}(\dots \text{swap}(\text{swap}(X_\ell^0, X_\ell^1, b_\ell^0), b_\ell^1), \dots, b_\ell^d). \quad (13)$$

for every  $\ell \in [1, m]$ . At the end of the inner loop, we hence correctly get (a randomized version of) the variable  $X_\ell^{b_\ell}$  copied in  $X_\ell^0$  for every  $\ell$ .

We shall consider that the  $\text{swap}$  procedure is composed of 2 algebraic operations (copies of algebraic variables) although the above description makes appear 4 copies. Indeed, in practice one could avoid having the same variables in input and output of the function and simply implement it as  $(Y^0, Y^1) \leftarrow (X^{0 \oplus b}, X^{1 \oplus b})$ . The total number  $N_\Pi$  of operations in the protected algorithm is hence given by:

$$N_\Pi = N_{pre} + N_{post} + n \cdot (N_{seq} + 5m(d+1)). \quad (14)$$

**Algorithm 4** Generic countermeasure

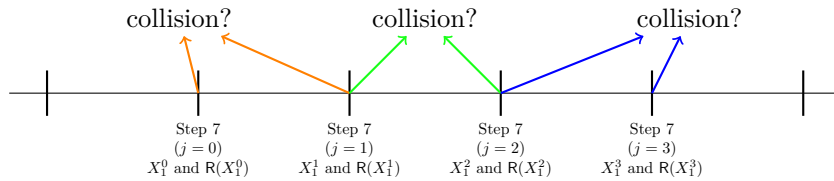
---

**Input:**  $\mathbf{X}, \widehat{\mathbf{k}} = ((k_0^j)_{0 \leq j \leq d}, (k_1^j)_{0 \leq j \leq d}, \dots, (k_{n-1}^j)_{0 \leq j \leq d})$   
**Output:**  $\mathbf{Y} = f_{\mathbf{k}}(\mathbf{X})$

1.  $(X_1^0, X_2^0, \dots, X_m^0) \leftarrow \text{preprocess}(\mathbf{X})$
2. **for**  $i = 0$  **to**  $n - 1$  **do**
3.  $((X_1^i, X_1^{i+1}), (X_2^i, X_2^{i+1}), \dots, (X_m^i, X_m^{i+1})) \leftarrow \text{sequence}(X_1^0, X_2^0, \dots, X_m^0)$
4. **for**  $j = 0$  **to**  $d$  **do**
5.  $(b_1^j, b_2^j, \dots, b_m^j) \leftarrow \text{indexes}(k_i^j)$
6. **for**  $\ell = 1$  **to**  $m$  **do**
7.  $X_\ell^0 \leftarrow \mathbf{R}(X_\ell^0)$
8.  $X_\ell^1 \leftarrow \mathbf{R}(X_\ell^1)$
9.  $((X_\ell^0, X_\ell^1)) \leftarrow \text{swap}(X_\ell^0, X_\ell^1, b_\ell^j)$
10. **end for**
11. **end for**
12. **end for**
13.  $\mathbf{Y} \leftarrow \text{postprocess}(X_1^0, X_2^0, \dots, X_m^0)$
14. **return**  $\mathbf{Y}$

---

**Practical security.** Our algorithm defeats advanced (higher-order) side-channel attacks. In particular, it is secure against template attacks of order  $d$  which target (masked) scalar manipulations (and related addresses), as well as against collision-based horizontal attacks of order  $d$ . The former attacks consist in targeting the direct manipulation of the sensitive bits (or digits)  $k_i$  or any address (or index in our formalism) that depends on  $k_i$ . In our countermeasure, all these sensitive bits and associated addresses are masked at order  $d$  and one must at least target a  $(d + 1)$ -tuple of such index variables to recover sensitive information (this is formally proved hereafter). The latter attacks consist in finding a collision between processed algebraic variables in order to infer information about the sensitive indexes. By our re-randomization and swap strategy, it can be checked that one must target at least  $d + 1$  algebraic variables to recover sensitive information. For instance, observing a collision between the variable  $X_\ell^0$  output of the randomization at Step 7 and the variable  $X_\ell^0$  output from the swap at Step 9 would allow one to learn that  $b_\ell^j = 0$ , while a non-collision would mean  $b_\ell^j = 1$ . A collision-based horizontal attacker should observe  $d + 1$  such (non-)collisions in order to learn the sensitive bit  $b_\ell$ . An illustration of such *higher-order* collision attack on Algorithm 4 is given in Figure 3 where the attacker observes successive executions of Step 7 to recover indexes  $b_1^j$ . Security against this kind of attacks is also captured by our formal proof.



**Figure 3:** Illustration of fourth-order collision attack on Algorithm 4.

**Practical efficiency.** From a high level viewpoint, our countermeasure to Algorithm 3 switches the complexity from  $\mathcal{O}(nm)$  to  $\mathcal{O}(nmd)$ , which means a linear slowdown with respect to the masking order  $d$ . However, we stress the practical efficiency of our generic countermeasure can be boosted in the case *lightweight* re-randomization technique can be applied. Indeed, our countermeasure does not add any algebraic operation to the original program besides copies and re-randomizations. In practice, the performances

of ECC and other exponentiation-based cryptosystems is mainly impacted by (field or ring) multiplications (or point doublings/additions at the group level), namely the operations that appear in the sequence procedure in Algorithm 3. If the cost of copying and re-randomization can be made negligible compared to costly multiplications then our countermeasure would be virtually free. Such a paradigm would be unrealistic for large values of the masking order  $d$  but the trade-off between costly algebraic operations and light copies and re-randomizations could lead to a reasonable impact on the performances.

We also note that our countermeasure is greedy in terms of randomness generation with the frequent use of re-randomization. But this is also true for other classical countermeasures achieving a high level of provable security in the noisy leakage model [DDF14, GJR18, DFZ19]. Moreover, this might be mitigated in practice by reducing the number of random bits used in a re-randomization. One could for instance argue that for the randomization in Example 1 the fresh random value  $R$  could be sampled from a smaller space than  $[0, h)$  (*i.e.* with less than  $\log_2(h)$  bits), which would constitute a trade-off between the hiddenness property and the efficiency. Let us stress once again that to the best of our knowledge and besides [ADP<sup>+</sup>20] (which shows the difficulty of defeating elliptic-curve point randomization), no attack has been published that defeats these types of randomization besides looking at the collision behavior of randomized variables (which is fully captured by our hiddenness notion with  $\varepsilon = 0$ ).

## 4.2 Security Proof

We state our formal security result in the following theorem.

**Theorem 1.** *The program  $\Pi$  defined in Algorithm 4 is  $\gamma$ -leakage resilient against  $\varepsilon$ -hiding  $\delta$ -noisy leakage functions with*

$$\gamma \leq n(5m\delta)^{d+1} + N_{\Pi} \cdot \varepsilon . \quad (15)$$

Theorem 1 shows that if  $\varepsilon$  is negligible and if  $\delta < \frac{1}{5m}$ , then our countermeasure achieves exponential security in the masking order  $d$ , at the cost of a linear slowdown for the protected algorithm.

*Remark 9.* In the above theorem, we implicitly assume that  $\delta$ -noisy leakage functions are defined with respect to the relative error distance (see (3)), as suggested in [PGMP19]. We stress that a different choice would not affect our proof except for the tightness of the bound (15). For instance, taking the  $L_1$ -norm statistical distance considered in [DDF14] would imply a factor  $|\mathcal{X}|$  to  $\delta$  in (15) (where  $\mathcal{X}$  is the domain of the leakage function). But as discussed in [PGMP19], this factor seems to be a proof artifact which does not apply while considering practically relevant leakage models.

We demonstrate Theorem 1 through a sequence of three games. Each game is composed of two experiments: a *real* experiment that outputs the real leakage distribution following the considered family of  $\delta$ -noisy leakage functions  $\mathcal{F}$ , and a *simulation* experiment that output the simulated leakage distribution. The  $\gamma$ -leakage resilience is achieved if and only if the output distributions of the two experiments are  $\gamma$ -close.

Game 0 (illustrated in Figure 4) exactly fits the definition of leakage resilience (Definition 3) where  $\mathcal{F}$  is a family of  $\varepsilon$ -hiding  $\delta$ -noisy leakage functions. We first show that this game reduces to Game 1 in which the `ComputTrace` algorithm is replaced by a different `ComputTrace'` algorithm. Then we show that Game 1 reduces to Game 2 in which the functions from  $\mathcal{F}$  are replaced by corresponding  $\delta$ -random probing functions. We finally exhibit a simulator which achieves  $\gamma_2$ -closeness of the experiment outputs in Game 2 for some  $\gamma_2$ . According to the two previous game transitions, we obtain  $\gamma_1$ -closeness and  $\gamma_0$ -closeness for the experiment outputs in Game 1 and Game 0 for some  $\gamma_1$  and  $\gamma_0$  that we shall exhibit throughout the proof.

ExpReal<sub>0</sub>( $\mathcal{A}, \Pi, \mathcal{F}$ ):

1.  $(\mathbf{X}, \mathbf{k}) \leftarrow \mathcal{A}()$
2.  $\widehat{\mathbf{k}} \leftarrow \text{Enc}(\mathbf{k})$
3.  $\mathcal{T} \leftarrow \text{ComputeTrace}(\Pi, \mathbf{X}, \widehat{\mathbf{k}})$
4.  $(\ell_1, \dots, \ell_z) \leftarrow \text{LeakageSampler}(\mathcal{T}, \mathcal{F})$
5. Return  $(\ell_1, \dots, \ell_z)$

ExpSim<sub>0</sub>( $\mathcal{A}, \Pi, \mathcal{F}, \text{Sim}$ ):

1.  $(\mathbf{X}, \mathbf{k}) \leftarrow \mathcal{A}()$
2.  $(\ell_1, \dots, \ell_z) \leftarrow \text{Sim}(\Pi, \mathbf{X}, \mathcal{F})$
3. Return  $(\ell_1, \dots, \ell_z)$

**Figure 4:** Game 0 (initial security game). The choice of the input  $(\mathbf{X}, \mathbf{k})$  is modeled by a probabilistic algorithm  $\mathcal{A}()$  (the adversary). The secret index variable  $\mathbf{k}$  is then encoded into  $\widehat{\mathbf{k}}$  using the Boolean sharing described in (9). Then the real experiment generates the computation trace  $\mathcal{T}$  for the program  $\Pi$  on input  $(\mathbf{X}, \widehat{\mathbf{k}})$  from which the leakage sampler produces the output  $(\ell_1, \dots, \ell_z)$  by applying the leakage functions from  $\mathcal{F}$ . On the other hand, the simulation experiment calls the simulator  $\text{Sim}$  on input  $\Pi, \mathbf{X}$  and  $\mathcal{F}$  which produces a simulated leakage output  $(\ell_1, \dots, \ell_z)$ .

ExpReal<sub>1</sub>( $\mathcal{A}, \Pi, \mathcal{F}$ ):

1.  $(\mathbf{X}, \mathbf{k}) \leftarrow \mathcal{A}()$
2.  $\widehat{\mathbf{k}} \leftarrow \text{Enc}(\mathbf{k})$
3.  $\mathcal{T} \leftarrow \text{ComputeTrace}'(\Pi, \mathbf{X}, \widehat{\mathbf{k}})$
4.  $(\ell_1, \dots, \ell_z) \leftarrow \text{LeakageSampler}(\mathcal{T}, \mathcal{F})$
5. Return  $(\ell_1, \dots, \ell_z)$

ExpSim<sub>1</sub>( $\mathcal{A}, \Pi, \mathcal{F}, \text{Sim}$ ):

1.  $(\mathbf{X}, \mathbf{k}) \leftarrow \mathcal{A}()$
2.  $(\ell_1, \dots, \ell_z) \leftarrow \text{Sim}(\Pi, \mathbf{X}, \mathcal{F})$
3. Return  $(\ell_1, \dots, \ell_z)$

**Figure 5:** Game 1. This game is similar to Game 0 but the (deterministic)  $\text{ComputeTrace}$  algorithm is replaced by a (probabilistic)  $\text{ComputeTrace}'$  algorithm. In the latter, the computation trace is produced  $\mathcal{T}$  is produced in a similar way as with the former algorithm, that is by evaluating the program on the inputs  $\mathbf{X}$  and  $\widehat{\mathbf{k}}$ . But this time, the output of each randomization operation  $R$  is randomly sampled from  $\mathcal{A}$  (according to the hiddenness property of the leakage).

ExpReal<sub>2</sub>( $\mathcal{A}, \Pi, \mathcal{F}'$ ):

1.  $(\mathbf{X}, \mathbf{k}) \leftarrow \mathcal{A}()$
2.  $\widehat{\mathbf{k}} \leftarrow \text{Enc}(\mathbf{k})$
3.  $\mathcal{T} \leftarrow \text{ComputeTrace}'(\Pi, \mathbf{X}, \widehat{\mathbf{k}})$
4.  $(\ell_1, \dots, \ell_z) \leftarrow \text{LeakageSampler}(\mathcal{T}, \mathcal{F}')$
5. Return  $(\ell_1, \dots, \ell_z)$

ExpSim<sub>2</sub>( $\mathcal{A}, \Pi, \mathcal{F}', \text{Sim}$ ):

1.  $(\mathbf{X}, \mathbf{k}) \leftarrow \mathcal{A}()$
2.  $(\ell_1, \dots, \ell_{z'}) \leftarrow \text{Sim}(\Pi, \mathbf{X}, \mathcal{F}')$
3. Return  $(\ell_1, \dots, \ell_z)$

**Figure 6:** Game 2. This game is similar to Game 1 but the leakage function family  $\mathcal{F}$  is replaced by a family  $\mathcal{F}'$  of  $\delta$ -random probing functions (formally defined hereafter in Definition 6). Specifically, for each  $\delta$ -noisy function  $f_i \in \mathcal{F}$ , there exists a  $\delta$ -random probing function  $f'_i$  such that  $f_i(\cdot) = g_i \circ f'_i(\cdot)$  for some noisy function  $g_i$  (see details in the proof of Lemma 2). The functions  $f_i$  in  $\mathcal{F}$  are replaced by those  $f'_i$  functions to build  $\mathcal{F}'$ .

Our first technical lemma (transition from Game 0 to Game 1) is given hereafter, and its proof is available in Appendix A.

**Lemma 1.** *Assume that there exists a simulator  $\text{Sim}$  such that  $\text{ExpReal}_1(\mathcal{A}, \Pi, \mathcal{F})$  and  $\text{ExpSim}_1(\mathcal{A}, \Pi, \mathcal{F}, \text{Sim})$  outputs  $\gamma_1$ -close distributions. Then, for the same simulator,  $\text{ExpReal}_0(\mathcal{A}, \Pi, \mathcal{F})$  and  $\text{ExpSim}_0(\mathcal{A}, \Pi, \mathcal{F}, \text{Sim})$  outputs  $\gamma_0$ -close distributions with*

$$\gamma_0 \leq \gamma_1 + N_{\Pi} \cdot \varepsilon \quad (16)$$

Before introducing our second technical lemma, we shall recall the definition of a random probing function.

**Definition 6** (Random probing function). A  $\delta$ -random probing function  $\varphi$  on a set  $\mathcal{X}$  is a noisy function which satisfies

$$\varphi(x) = \begin{cases} x & \text{with probability } \delta \\ \perp & \text{with probability } 1 - \delta \end{cases} \quad (17)$$

for every  $x \in \mathcal{X}$ , where  $\perp$  denotes a void output and  $\text{Im}(\varphi) = \mathcal{X} \cup \{\perp\}$ .

In the above definition, we skip the randomness argument from the definition of the function  $\varphi$  as discussed in [Remark 2](#). In the context of a random probing function, this randomness determines whether the event  $\varphi(\cdot) = \perp$  occurs (with probability  $\delta$ ) or not. We can now formally state our second technical lemma (transition from Game 1 to Game 2) whose proof is in [Appendix A](#).

**Lemma 2.** *For every family  $\mathcal{F}$  of  $\delta$ -noisy functions, there exists a family  $\mathcal{F}'$  of  $\delta$ -random probing functions, for which the following holds.*

- if there exists a simulator  $\text{Sim}'$  such that the experiments  $\text{ExpReal}_2(\mathcal{A}, \Pi, \mathcal{F}')$  and  $\text{ExpSim}_2(\mathcal{A}, \Pi, \mathcal{F}', \text{Sim}')$  outputs  $\gamma_2$ -close distributions,
- then there exists a simulator  $\text{Sim}$  such that the experiments  $\text{ExpReal}_1(\mathcal{A}, \Pi, \mathcal{F})$  and  $\text{ExpSim}_1(\mathcal{A}, \Pi, \mathcal{F}, \text{Sim})$  outputs  $\gamma_1$ -close distributions with  $\gamma_1 \leq \gamma_2$ .

We finally give our last technical lemma on the security of Game 2. Its proof is available in [Appendix A](#).

**Lemma 3.** *Let  $\Pi$  be the randomized RAP described in [Algorithm 4](#) for masking order  $d$  and let  $\mathcal{F}'$  be a family of  $\delta$ -random probing functions. There exists a simulator  $\text{Sim}$  such that  $\text{ExpReal}_2(\mathcal{A}, P, \mathcal{F}')$  and  $\text{ExpSim}_2(\mathcal{A}, P, \mathcal{F}', \text{Sim})$  output  $\gamma_2$ -close distributions, with*

$$\gamma_2 \leq n(5m\delta)^{d+1} . \quad (18)$$

## 5 Applications

We now provide concrete examples of regular algorithms – including examples from [Section 2](#) – and show that they perfectly fit the generic shape of [Algorithm 3](#). We first show how to apply our generic countermeasure to various binary scalar multiplication algorithms, and then to the classic left-to-right fixed-window algorithm (see [Algorithm 2](#)).

### Application to Binary Ladders

**Montgomery ladder.** To rewrite the Montgomery ladder (see [Algorithm 1](#)) in our framework, observe that a loop iteration can be rewritten as:

$$\left\{ \begin{array}{l} b \leftarrow k_i \\ (R_1, R_0) \leftarrow (R_{1-b}, R_b) \\ R_1 \leftarrow R_1 + R_0 \\ R_0 \leftarrow 2 \cdot R_0 \\ (R_1, R_0) \leftarrow (R_{1-b}, R_b) \end{array} \right. \iff \left\{ \begin{array}{l} b \leftarrow k_{i+1} \oplus k_i \\ (R_1, R_0) \leftarrow (R_{1-b}, R_b) \\ R_1 \leftarrow R_1 + R_0 \\ R_0 \leftarrow 2 \cdot R_0 \end{array} \right.$$

where the above equivalence holds by defining  $k_n := 0$  and returning  $R_{k_0}$  instead of  $R_0$  in the latter case. We thus obtain [Algorithm 5](#) which rewrites the Montgomery ladder in the framework of [Algorithm 3](#), where at the end of each iteration  $(X_1^0, X_2^0)$  matches  $(R_1, R_0)$ , while  $(X_1^1, X_2^1)$  matches  $(R_0, R_1)$ .

**Algorithm 5** Montgomery ladder at the point level**Input:**  $P, k = (k_n, k_{n-1}, \dots, k_1, k_0)$  with  $k_n := 0$ **Output:**  $[k]P$ 

1.  $(X_1^0, X_2^0) \leftarrow (P, \mathcal{O}); (X_1^1, X_2^1) \leftarrow (\mathcal{O}, P)$  // preprocess( $P$ )
2. **for**  $i = 0$  **to**  $n - 1$  **do**
3.    $b \leftarrow k_{i+1} \oplus k_i$  // indexes( $k_i$ )
4.    $(X_1^0, X_2^0) \leftarrow (X_1^b, X_2^b)$
5.    $(X_1^0, X_2^0) \leftarrow (X_1^0 + X_2^0, 2X_2^0)$  // sequence( $X_1^0, X_2^0$ )
6.    $(X_1^1, X_2^1) \leftarrow (X_2^0, X_1^0)$
7. **end for**
8.  $b \leftarrow k_0$  // indexes( $k_i$ )
9.  $(X_1^0, X_2^0) \leftarrow (X_1^b, X_2^b)$
10. **return**  $X_2^0$

**Montgomery ladder at coordinate level.** Our approach can also be applied at the coordinate level *i.e.* when internal registers store field element rather than points. We illustrate this on the Montgomery ladder recently exhibited by Hamburg [Ham20] and which is today the most efficient binary ladder for standard Weierstrass curves. Algorithm 6 rewrites this algorithm in our framework. The interested reader is referred to the original paper [Ham20] for the descriptions of `setup-ladder` and `finalize-ladder` functions). For more efficiency, the computations from lines 7 to 11 have been deported after the selection of the algebraic variables (line 6). As for our countermeasure, they can be seen as belonging to the `sequence` operation of the next iteration. Also, we consider here that the scalar has been rewritten so that the indexes can be selected from each bit directly.

**Algorithm 6** Montgomery ladder at the coordinate level from [Ham20]**Input:**  $P, k = (k_{n-1}, \dots, k_1, k_0)$ **Output:**  $[k]P$ 

1.  $(X_{QP}, X_{RP}, M, Y_P) \leftarrow \text{setup-ladder}(P)$
2.  $(X_1^0, X_2^0, X_3^0, X_4^0) \leftarrow (X_{QP}, X_{RP}, M, Y_P)$  // preprocess( $P$ )
3. **for**  $i = 0$  **to**  $n - 1$  **do**
4.    $(X_1^0, X_1^1), (X_2^0, X_2^1) \leftarrow (X_2^0, X_1^0), (X_1^0 - X_2^0, X_2^0 - X_1^0)$  // sequence( $X_1^0, X_2^0$ )
5.    $(b_1, b_2) \leftarrow (k_i, k_i)$  // indexes( $k_i$ )
6.    $(X_1^0, X_2^0) \leftarrow (X_1^{b_1}, X_2^{b_2})$
7.    $T \leftarrow Y_P + 2 \cdot M \cdot X_1^0$
8.    $X_1^0 \leftarrow T^2 \cdot (T^2 + 2 \cdot M \cdot T \cdot X_2^0)$
9.    $X_2^0 \leftarrow (X_1^0 \cdot (X_2^0)^2) + Y_P \cdot T \cdot (X_2^0)^3$
10.    $X_3^0 \leftarrow X_1^0 (X_2^0)^2 - T^2 - M \cdot T \cdot X_2^0$
11.    $X_4^0 \leftarrow Y_P \cdot T^3 \cdot (X_2^0)^3$
12. **end for**
13.  $Y \leftarrow \text{finalize-ladder}(X_1^0, X_2^0, X_3^0, X_4^0)$  // postprocess( $X_1^0, X_2^0, X_3^0, X_4^0$ )
14. **return**  $Y$

We note that the Montgomery ladder for Curve25519 and Curve448 described in RFC 7748 [LHT16] at the coordinate level is already expressed in a similar paradigm as ours (in particular making use of conditional swap instructions). Our generic countermeasure therefore directly applies to this algorithm.

**Joye ladder.** Similarly, Algorithm 7 rewrites the double-and-add algorithm of Joye [Joy07] so that it perfectly matches our generic regular program displayed in Algorithm 3.



**Algorithm 7** Joye ladder

---

**Input:**  $P, k = (k_0, k_1, \dots, k_{n-1})$   
**Output:**  $Y = f_k(P)$

1.  $(X_1^0, X_2^0) \leftarrow (\mathcal{O}, P)$  // preprocess( $P$ )
2. **for**  $i = 0$  **to**  $n - 1$  **do**
3.    $T \leftarrow 2X_1^0 + X_2^0$  // sequence( $X_1^0, X_2^0$ )
4.    $(X_1^0, X_1^1), (X_2^0, X_2^1) \leftarrow (T, X_1^0), (T, X_2^0)$
5.    $(b_1, b_2) \leftarrow (1 - k_i, k_i)$  // indexes( $k_i$ )
6.    $(X_1^0, X_2^0) \leftarrow (X_1^{b_1}, X_2^{b_2})$
7. **end for**
8. **return**  $X_1^0$

---

**Signed binary ladder.** Another example is the signed binary algorithm from [Riv11]. It is particularly efficient when the points are represented using  $(X, Y)$ -only co- $Z$  coordinates [GJM<sup>+</sup>11, Riv11]. In this algorithm, the input scalar is assumed to satisfy  $k_0 = k_{n-1} = 1$  and it is encoded under its (unique) full signed binary expansion in which each digit belongs to  $\{-1, 1\}$ . Applying the standard left-to-right algorithm to this signed expansion yields a regular signed binary ladder which can also be rewritten in our framework as shown in Algorithm 8.

**Algorithm 8** Signed binary ladder [Riv11, Alg. 8]

---

**Input:**  $P, k' = (k_{n-1} \oplus k_{n-2}, \dots, k_2 \oplus k_1, k_1 \oplus 1)_2$   
**Output:**  $[k]P$

1.  $(X_1^0, X_2^0) \leftarrow (P, 3P)$  // preprocess( $P$ )
2. **for**  $i = 0$  **to**  $n - 2$  **do**
3.    $T \leftarrow 2X_1^0 + X_2^0$  // sequence( $X_1^0, X_2^0$ )
4.    $(X_1^0, X_1^1), (X_2^0, X_2^1) \leftarrow (T, T), (-X_2^0, X_2^0)$
5.    $b \leftarrow 1 - k'_i$  // indexes( $k_i$ )
6.    $(X_1^0, X_2^0) \leftarrow (X_1^b, X_2^b)$
7. **end for**
8.  $Y \leftarrow X_1^0$  // postprocess( $X_1^0, X_2^0$ )
9. **return**  $Y$

---

From the rewrites of the binary ladders, we can directly apply our generic countermeasure and obtain a leakage resilient program. In a nutshell, for some masking order  $d$ , the steps in Algorithm 5, Algorithm 7, and Algorithm 8 that correspond to the `sequence` procedure will remain unchanged, while registers  $(X_1^0, X_1^1)$  and  $(X_2^0, X_2^1)$  will be refreshed  $d + 1$  times and pairwise swapped in place of the scalar-dependent selection step. The obtained leakage resilient program achieves the provable security result of Theorem 1 with  $m = 2$ .

**Application to Fixed-Window Scalar Multiplication** The fixed-window scalar multiplication described in Algorithm 2 can also be rewritten in the generic shape of Algorithm 3, at the cost of a few differences (see Algorithm 9). As explained in Section 2, this algorithm requires the pre-computation of a table  $T$  to store the multiples  $dP$  for digits  $d$  belonging to some basis  $\mathbb{B} \subset \mathbb{Z}$ . Different choices are possible for such a basis depending on the context. Our description of Algorithm 9 implicitly assumes  $|\mathbb{B}| = 2^w$ . For any digit  $d \in \mathbb{B}$ , we denote  $\tilde{d} \in [0, 2^w - 1]$  its index in the basis  $\mathbb{B}$ , and assume that each digit  $k_i$  of the scalar is represented by its  $w$ -bit index  $\tilde{k}_i$  (the sharing being applied to those index values).

Contrary to binary algorithms, Algorithm 9 loops on  $w$ -bit windows and it requires  $2^w$  temporary registers  $(X^i)_{0 \leq i < 2^w - 1}$  instead of the two registers per input introduced in Algorithm 3. Finally, to ease the exposition we consider two `sequence` procedures, at the

beginning and at the end of the loop iterations. Note that the `sequence` procedure at the end of an iteration could be considered to be part of the next iteration to fit to the original model (and for the last iteration, it can be seen as a sub-step of the post processing). In both cases, this step only counts in the number of operations of each of these generic functions, but does not alter the security proof.

---

**Algorithm 9** Fixed window
 

---

**Input:**  $P$ ,  $\mathbf{k} = (k_{n-1}, \dots, k_1, k_0)_{2^w}$ 
**Output:**  $[k]P$ 

```

1. forall  $d \in \mathbb{B}$  do  $T[\tilde{d}] = [d]P$  // preprocess( $P$ )
2. for  $i = 0$  to  $n - 1$  do
3.    $(X^0, X^1, \dots, X^{2^w-1}) \leftarrow T$  // sequence1( $X$ )
4.    $c \leftarrow \tilde{k}_i$  // indexes( $k_i$ )
5.    $X^0 \leftarrow X^c$ 
6.    $X \leftarrow X^0 + [2^w]X$  // sequence2( $X$ )
7. end for
8.  $Y \leftarrow X$  // postprocess( $X^0$ )
9. return  $Y$ 

```

---

The main difference while applying our countermeasure is in the type of the index  $c$  at Step 6 which is not a bit anymore but a  $w$ -bit coordinate of  $\mathbf{k}$ . The swapping procedure must hence be generalized to work with  $w$ -bit Boolean shares, which can be done by defining it as:

$$\text{swap}(X^0, X^1, \dots, X^{2^w-1}, c) = (X^{c \oplus 0}, X^{c \oplus 1}, \dots, X^{c \oplus (2^w-1)}). \quad (19)$$

Under such a definition, we still have the desired property

$$\text{swap}(X^0, \dots, X^{2^w-1}, c) = \text{swap}(\dots \text{swap}(\text{swap}(X^0, \dots, X^{2^w-1}, c^0), c^1), \dots, c^d)$$

which is satisfied for every Boolean sharing  $c^0 \oplus c^1 \oplus \dots \oplus c^d = c$ . While applying our countermeasure, Step 6 can thus be replaced by:

```

for  $j = 0$  to  $d$  do
   $c^j \leftarrow \tilde{k}_i^j$ 
   $(X^0, X^1, \dots, X^{2^w-1}) \leftarrow (R(X^0), R(X^1), \dots, R(X^{2^w-1}))$ 
   $(X^0, X^1, \dots, X^{2^w-1}) \leftarrow \text{swap}(X^0, X^1, \dots, X^{2^w-1}, c)$ 
end for

```

Our security proof works the same way on the obtained program which hence satisfies the leakage-resilience result of [Theorem 1](#).

## 6 Performance

### 6.1 Performance Analysis

A high-level complexity analysis of our countermeasure was provided in [Section 4](#). Namely, we reach a complexity in  $\mathcal{O}(nmd)$  with  $n$  the scalar bit length,  $m$  the number of cells of the algebraic memory (*i.e.*, the number of algebraic variables), and  $d$  the masking order. Going a little deeper into the details, denoting  $C_{\text{op}}$  the complexity of operation `op`, the overall complexity of [Alg. 4](#) is exactly equal to

$$C_{\text{preprocess}} + C_{\text{postprocess}} + n \cdot (C_{\text{sequence}} + (d + 1) \cdot (C_{\text{indexes}} + m \cdot (2 \cdot C_{\text{R}} + C_{\text{swap}}))).$$

The complexity of the `indexes` operation  $C_{\text{indexes}}$  being reduced to  $m$  index operations and the complexity of the `swap` operation to two algebraic copies, the prominent operations

remain the multiplications in **sequence** and the randomization of algebraic variables  $R$ . We estimate the randomization complexity  $C_R$  when implemented with one of the two examples displayed in Section 3, and also with both. We shall further denote  $C_{\text{mult}_\alpha}$  the complexity of a multiplication in  $\mathbb{Z}_q$  between operands of bit sizes  $\alpha$  and approximate its cost with  $c_m \cdot \alpha^2$  for some constant  $c_m$ .

In Example 1, we consider the randomization operation which maps an element of  $\mathbb{Z}_p$  to a randomized representation in  $\mathbb{Z}_{hp}$ , namely  $R(x) = x + R \cdot p$  with  $R \leftarrow [0, h)$ . It requires a multiplication in  $\mathbb{Z}_{hp}$  of complexity

$$C_{R_1} = c_m \cdot |p| \cdot |h|$$

and an addition in  $\mathbb{Z}_{hp}$  that we shall consider to be negligible compared to the multiplication. Using this randomization in Alg. 4 would require the **sequence** operation and all the subsequent randomizations to be performed in  $\mathbb{Z}_{hp}$ . Denoting by  $\gamma$  the number of multiplications in **sequence**, the overall complexity would remain equivalent to

$$n \cdot (\gamma \cdot C_{\text{mult}_{|p|+|h|}} + 2 \cdot (d+1) \cdot m \cdot C_{R_1})$$

In Example 2, the randomization operation maps a point defined by its Jacobian coordinates on an elliptic curve of base field  $\mathbb{Z}_p$  to a randomized representation, namely  $R(\mathbf{P}) = R((X : Y : Z)) = (X\lambda^2 : Y\lambda^3 : Z\lambda)$  with  $\lambda \leftarrow \mathbb{Z}_p^*$ . This randomization mainly requires 5 multiplications between elements in  $\mathbb{Z}_p$  or  $\mathbb{Z}_p^*$  (assuming that the squaring operation  $\lambda^2$  is performed by a multiplication). We thus express its complexity as follows:

$$C_{R_2} = 5 \cdot C_{\text{mult}_{|p|}} = 5 \cdot c_m \cdot |p|^2.$$

This complexity could be reduced in our countermeasure when considering the manipulation of two points in co- $Z$  coordinates in the main loop. The double randomization of complexity  $2 \cdot C_{R_2}$  could then be performed with only 6 multiplications. This would result in an overall complexity equivalent to

$$n \cdot (\gamma \cdot C_{\text{mult}_{|p|}} + 6 \cdot (d+1) \cdot m \cdot C_{\text{mult}_{|p|}})$$

One step further, we could gather both randomization techniques to improve the security level (*i.e.*, lower parameter  $\varepsilon$ ) of our countermeasure. We would randomize the Jacobian coordinates, previously augmented with multiples of the modulus  $p$ . With co- $Z$  coordinates, the overall complexity would thus be equivalent to

$$n \cdot (\gamma \cdot C_{\text{mult}_{|p|+|h|}} + 10 \cdot (d+1) \cdot m \cdot C_{\text{mult}_{|p|+|h|}}).$$

Table 1 illustrates our countermeasure's complexity by giving the multiplicative factor between a regular program and a program with one randomization or the other at different masking orders. Neglecting the linear operations, we evaluate the overall complexity as an equivalent of

$$n \cdot (C_{\text{sequence}} + (d+1) \cdot (C_{\text{indexes}} + m \cdot (2 \cdot C_R)))$$

for protected operations and

$$n \cdot C_{\text{sequence}}$$

for regular programs. Following the example of the Montgomery ladder in Algorithm 5, we assume one point addition and one point doubling in the **sequence** step, *i.e.* around 12 multiplications. We also assume that we have 2 algebraic variables and  $h$  can vary in  $\{32, 64, 128\}$ .  $\lambda$  for the second randomization is generated in the field  $(\mathbb{Z}_p$  or  $\mathbb{Z}_{hp})$ . We also take  $p$  of size 256 bits.

**Comparison with naive countermeasures.** For scalar multiplications for instance, a first alternative countermeasure would be to split the scalar into  $d + 1$  shares and compute all the instructions of Algorithm 3 ( $d + 1$ ) times to obtain  $d + 1$  additive shares of the result. This method would suffer from two main issues. Regarding the practical security of the countermeasure, if the original computation was vulnerable to a one-trace attack (horizontal and/or collision attack), then an attacker would easily break the proposed scheme by running  $d + 1$  one-trace attacks on each computation to retrieve the  $d + 1$  shares. Conversely, our proposal thwarts such one-trace attacks through regular re-randomization. On the theoretical side, a formal proof of the simple scalar splitting scheme in our model would demonstrate a result similar to Theorem 1, but instead of having  $(5m\delta)$  to the power  $d + 1$ , one would have a much greater term than  $5m$  (basically including the number of algebraic operations in the whole scalar multiplication) which would imply a much stronger requirement on  $\delta$  (*i.e.* on the amount of noise). Regarding the complexity, the heavier operations (*e.g.* multiplications) of Algorithm 3 are performed in the **sequence** operation which is independent from the number of shares.<sup>5</sup> In contrast, the simple scalar splitting scheme would imply increasing the number of multiplications by a factor  $(d + 1)$ . We see from Table 1 that, besides achieving tighter security (under the hiddenness assumption), our countermeasure with  $R_1$  outperforms the scalar splitting countermeasure. For the considered example (with  $\gamma = 12$  and  $m = 4$ ), our countermeasure with  $R_2$  induces a slightly higher overhead than scalar splitting. We stress that the cost of  $R_2$  could also be mitigated by taking a smaller random  $\lambda$ , of *e.g.* 32, 64 or 128 bits, instead of a random element of  $\mathbb{Z}_p$  (256 bits).

The last row of the Table 1 estimates the complexity of a second naive approach which consists in applying the Ishai-Sahai-Wagner (ISW) scheme to protect the scalar multiplication circuit [ISW03]. The latter would notably require  $(d+1)^2$  multiplications per multiplication in the original circuit. Note that Table 1 only counts the overhead in terms of multiplication, which is optimistic since displays the ISW multiplications would not be enough to achieve probing security (further protections would be additionally required). We see that our countermeasure quickly outperforms this approach as  $d$  grows. Furthermore, such a probing secure countermeasure would not protect ECC implementations against horizontal attacks which are not captured by this leakage model [BCPZ16].

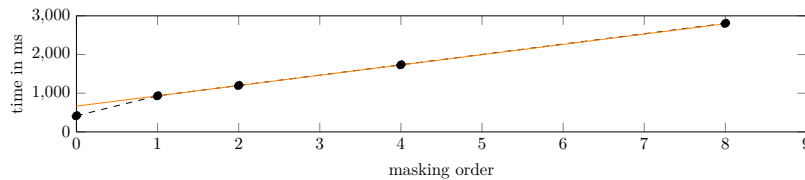
**Table 1:** Complexity Approximation

	order 1	order 2	order 4	order 8
Our countermeasure (overhead)				
$R_1 - h = 32$	1,35	1,39	1,47	1,64
$R_1 - h = 64$	1,73	1,81	1,98	2,31
$R_1 - h = 128$	2,58	2,75	3,08	3,75
$R_2$	3	4	6	10
$R_1 \& R_2 - h = 32$	5,48	7,59	11,81	20,25
$R_1 \& R_2 - h = 64$	6,77	9,38	14,58	25
$R_1 \& R_2 - h = 128$	9,75	13,5	21	36
Other countermeasures (overhead)				
scalar splitting	2	3	5	9
naive ISW	4	9	25	81

<sup>5</sup>This is not well illustrated on our practical experiments (because the generation of randomness is very expensive on the considered device).

**Table 2:** Benchmarks of Alg. 8 and its countermeasure at several orders.

	Regular program	Our countermeasure			
		order 1	order 2	order 4	order 8
card randomness	416,5 ms	933,1 ms	1200,8 ms	1734,0 ms	2804,9 ms
free randomness	415,2 ms	612,4 ms	718,5 ms	930,4 ms	1355,9 ms
RAM complexity	1,220 kB	1,253 kB	1,286 kB	1,352 kB	1,484 kB

**Figure 7:** Benchmarks from Table 2 where masking order 0 corresponds to a regular algorithm. The orange line represents the function  $x \mapsto 400 + 266(x + 1)$ .

## 6.2 Implementation Benchmarks

We provide performance benchmarks for a concrete protected scalar multiplication implementation on a 0.13um 32-bit Contact Smartcard IC. The IC was EAL4+ certified in Asia and features an ARM SecureCore SC100 processor with 18KB of RAM, 8KB of ROM, and 548KB of FLASH.<sup>6</sup>

Our implementation is based on the signed binary ladder using  $(X, Y)$ -only co- $Z$  coordinates [Riv11, Algorithm 8].<sup>7</sup> In the basic setting (without applying our countermeasure), the scalar is randomized by the addition of a 288-bit multiple of the curve order and the Jacobian coordinates of the points are randomized before starting the main loop. Benchmarks for this implementation with the NIST elliptic curve P-256 [NIS13] are provided in the column *Regular program* of Figure 7 (averaged over 100 executions).

Then, we modified our implementation to follow the description of Alg. 8 and applied the generic countermeasure introduced in Section 4 for a customizable masking order. The randomization step follows the described randomization of Jacobian coordinates simultaneously applied on two points in  $(X, Y)$ -only coordinates as well as on additional inverted  $Y$ -coordinate. This makes a total of 2 multiplications to derive  $\lambda^2$  and  $\lambda^3$  from  $\lambda$  plus 5 multiplications with the point coordinates. The additional cost of this step is mainly due to these seven extra multiplications. We provide the corresponding benchmarks at orders 1, 2, 4, and 8. From these results, we can interpolate a running time of  $400 + 266(d + 1)$  ms for a masking order  $d$ , as displayed in Table 2 and in Figure 7. The performance of the first row are computed when the randomness is generated on the smart card. Using a constant instead (to simulate the cost of the randomness generation), we obtain the results displayed in the second row. The randomness cost in the regular program is very light (*i.e.*, the difference is low between the complexity when using the card and that when assuming the randomness is free) since, without our countermeasure, only a little randomness is actually used. Note that the gap between the regular program and the first-order masked implementation comes from the randomization which is applied directly on both shares in the latter case. As for the memory complexity, our regular program requires 1220 bytes of RAM while the countermeasure add 33 bytes per masking order.

<sup>6</sup>We refer the interested reader to [CCC<sup>+</sup>19] for further details on this device.

<sup>7</sup>The C source code of our implementation is available at <https://github.com/CryptoExperts/high-order-ecc-implementation>

## 7 Conclusion

In this paper, we have introduced the first formal model of computation to capture regular implementations of ECC and other exponentiation-based cryptosystems (*e.g.*, RSA, discrete logarithm-based schemes, pairings) in the presence of noisy leakage. We have further introduced the *hiddenness* property, which captures the effect of randomization operations on the leakage of algebraic variables in such protected implementations. The hiddenness assumption states that the leakage of a randomized variable is hard to distinguish from the leakage of a random value for some given leakage function  $f$ . We show that for practically relevant pairs of randomization operation and leakage function, this assumption is heuristically verified in practice, which is further confirmed by the absence of underlying attacks in the literature. From this ground, we design a generic countermeasure that can be applied to a wide range of regular ECC implementations and exponentiation-based cryptosystems. Our approach is to protect the scalar using high order Boolean masking and to protect the algebraic variables using a randomization technique which assumably achieves hiddenness security for the considered leakage. In order to obtain provable security against high order address-bit and/or collision attacks (and more generally provable high order side-channel security) our countermeasure relies on a many-time swapping and re-randomization approach. We provide a formal security proof of its security in the noisy leakage model under the hiddenness assumption. We further demonstrate the applicability of our generic countermeasure on several widely used scalar multiplication algorithms, provide a complexity analysis and comparison to other approaches and report concrete benchmarks for a smartcard implementation.

Our work paves the way to the design of efficient ECC implementations achieving high orders of provable side-channel security. Further research would be welcome to study the hiddenness assumption in practice for different randomization techniques. This question is related to the design of side-channel attacks that exploit the leakage of randomized variables to recover information on the unrandomized variables (beyond the collision behavior). This issue has not been widely investigated so far, presumably due to the hardness of exploiting such leakage. Assessing the hiddenness security for different leakage functions (on different hardware targets) and with respect to the size of the random elements used to re-randomize the algebraic variables would be of great interest to determine the best efficiency-security trade-offs for implementations protected in our framework.

**Acknowledgments.** This work was partly supported by the French FUI-AAP25 VeriSiCC project.

## References

- [ADP<sup>+</sup>20] Melissa Azouaoui, François Durvaux, Romain Poussier, François-Xavier Standaert, Kostas Papagiannopoulos, and Vincent Verneuil. On the worst-case side-channel security of ECC point randomization in embedded devices. In Karthikeyan Bhargavan, Elisabeth Oswald, and Manoj Prabhakaran, editors, *INDOCRYPT 2020*, volume 12578 of *LNCS*, pages 205–227. Springer, Heidelberg, December 2020.
- [BCO04] Eric Brier, Christophe Clavier, and Francis Olivier. Correlation power analysis with a leakage model. In Marc Joye and Jean-Jacques Quisquater, editors, *CHES 2004*, volume 3156 of *LNCS*, pages 16–29. Springer, Heidelberg, August 2004.

- [BCP<sup>+</sup>19] Lejla Batina, Lukasz Chmielewski, Louiza Papachristodoulou, Peter Schwabe, and Michael Tunstall. Online template attacks. *Journal of Cryptographic Engineering*, 9(1):21–36, April 2019.
- [BCPZ16] Alberto Battistello, Jean-Sébastien Coron, Emmanuel Prouff, and Rina Zeitoun. Horizontal side-channel attacks and countermeasures on the ISW masking scheme. In Benedikt Gierlichs and Axel Y. Poschmann, editors, *CHES 2016*, volume 9813 of *LNCS*, pages 23–39. Springer, Heidelberg, August 2016.
- [BGP<sup>+</sup>11] Lejla Batina, Benedikt Gierlichs, Emmanuel Prouff, Matthieu Rivain, François-Xavier Standaert, and Nicolas Veyrat-Charvillon. Mutual information analysis: a comprehensive study. *Journal of Cryptology*, 24(2):269–291, April 2011.
- [BJ02] Eric Brier and Marc Joye. Weierstraß elliptic curves and side-channel attacks. In David Naccache and Pascal Paillier, editors, *PKC 2002*, volume 2274 of *LNCS*, pages 335–345. Springer, Heidelberg, February 2002.
- [BJPW14] Aurélie Bauer, Éliane Jaulmes, Emmanuel Prouff, and Justine Wild. Horizontal collision correlation attack on elliptic curves. In Tanja Lange, Kristin Lauter, and Petr Lisonek, editors, *SAC 2013*, volume 8282 of *LNCS*, pages 553–570. Springer, Heidelberg, August 2014.
- [BL07a] Daniel J. Bernstein and Tanja Lange. Faster addition and doubling on elliptic curves. In Kaoru Kurosawa, editor, *ASIACRYPT 2007*, volume 4833 of *LNCS*, pages 29–50. Springer, Heidelberg, December 2007.
- [BL07b] Daniel J. Bernstein and Tanja Lange. Inverted edwards coordinates. In Serdar Boztas and Hsiao-feng Lu, editors, *AAECC-17*, volume 4851 of *LNCS*, pages 20–27. Springer, 2007.
- [BvSY14] Naomi Benger, Joop van de Pol, Nigel P. Smart, and Yuval Yarom. “ooh aah... just a little bit”: A small amount of side channel can go a long way. In Lejla Batina and Matthew Robshaw, editors, *CHES 2014*, volume 8731 of *LNCS*, pages 75–92. Springer, Heidelberg, September 2014.
- [CCC<sup>+</sup>19] Mathieu Carbone, Vincent Conin, Marie-Angela Cornélie, François Dassance, Guillaume Dufresne, Cécile Dumas, Emmanuel Prouff, and Alexandre Venelli. Deep learning to evaluate secure RSA implementations. *IACR TCHES*, 2019(2):132–161, 2019. <https://tches.iacr.org/index.php/TCHES/article/view/7388>.
- [CCJ04] Benoît Chevallier-Mames, Mathieu Ciet, and Marc Joye. Low-cost solutions for preventing simple side-channel analysis: Side-channel atomicity. *IEEE Trans. Computers*, 53(6):760–768, 2004.
- [CDP17] Eleonora Cagli, Cécile Dumas, and Emmanuel Prouff. Convolutional neural networks with data augmentation against jitter-based countermeasures - profiling attacks without pre-processing. In Wieland Fischer and Naofumi Homma, editors, *CHES 2017*, volume 10529 of *LNCS*, pages 45–68. Springer, Heidelberg, September 2017.
- [CFG<sup>+</sup>10] Christophe Clavier, Benoit Feix, Georges Gagnerot, Mylène Roussellet, and Vincent Verneuil. Horizontal correlation analysis on exponentiation. In Miguel Soriano, Sihang Qing, and Javier López, editors, *ICICS 10*, volume 6476 of *LNCS*, pages 46–61. Springer, Heidelberg, December 2010.

- [CJ01] Christophe Clavier and Marc Joye. Universal exponentiation algorithm. In Çetin Kaya Koç, David Naccache, and Christof Paar, editors, *CHES 2001*, volume 2162 of *LNCS*, pages 300–308. Springer, Heidelberg, May 2001.
- [Cor99] Jean-Sébastien Coron. Resistance against differential power analysis for elliptic curve cryptosystems. In Çetin Kaya Koç and Christof Paar, editors, *CHES'99*, volume 1717 of *LNCS*, pages 292–302. Springer, Heidelberg, August 1999.
- [CPRR14] Jean-Sébastien Coron, Emmanuel Prouff, Matthieu Rivain, and Thomas Roche. Higher-order side channel security and mask refreshing. In Shiho Moriai, editor, *FSE 2013*, volume 8424 of *LNCS*, pages 410–424. Springer, Heidelberg, March 2014.
- [CRR03] Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. Template attacks. In Burton S. Kaliski Jr., Çetin Kaya Koç, and Christof Paar, editors, *CHES 2002*, volume 2523 of *LNCS*, pages 13–28. Springer, Heidelberg, August 2003.
- [CT06] Thomas M. Cover and Joy A. Thomas. *Elements of information theory, Second Edition*. Wiley-Interscience, 2006.
- [DDF14] Alexandre Duc, Stefan Dziembowski, and Sebastian Faust. Unifying leakage models: From probing attacks to noisy leakage. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 423–440. Springer, Heidelberg, May 2014.
- [DFS16] Stefan Dziembowski, Sebastian Faust, and Maciej Skórski. Optimal amplification of noisy leakages. In Eyal Kushilevitz and Tal Malkin, editors, *TCC 2016-A, Part II*, volume 9563 of *LNCS*, pages 291–318. Springer, Heidelberg, January 2016.
- [DFZ19] Stefan Dziembowski, Sebastian Faust, and Karol Zebrowski. Simple refreshing in the noisy leakage model. In Steven D. Galbraith and Shiho Moriai, editors, *ASIACRYPT 2019, Part III*, volume 11923 of *LNCS*, pages 315–344. Springer, Heidelberg, December 2019.
- [DPN<sup>+</sup>16] Margaux Dugardin, Louiza Papachristodoulou, Zakaria Najm, Lejla Batina, Jean-Luc Danger, and Sylvain Guilley. Dismantling real-world ECC with horizontal and vertical template attacks. In François-Xavier Standaert and Elisabeth Oswald, editors, *COSADE 2016*, volume 9689 of *LNCS*, pages 88–108. Springer, Heidelberg, April 2016.
- [FGP<sup>+</sup>18] Sebastian Faust, Vincent Grosso, Santos Merino Del Pozo, Clara Paglialonga, and François-Xavier Standaert. Composable masking schemes in the presence of physical defaults & the robust probing model. *IACR TCHES*, 2018(3):89–120, 2018. <https://tches.iacr.org/index.php/TCHES/article/view/7270>.
- [FJ10] Reza Rezaeian Farashahi and Marc Joye. Efficient arithmetic on Hessian curves. In Phong Q. Nguyen and David Pointcheval, editors, *PKC 2010*, volume 6056 of *LNCS*, pages 243–260. Springer, Heidelberg, May 2010.
- [FLS14] Armando Faz-Hernández, Patrick Longa, and Ana H. Sanchez. Efficient and secure algorithms for GLV-based scalar multiplication and their implementation on GLV-GLS curves. In Josh Benaloh, editor, *CT-RSA 2014*, volume 8366 of *LNCS*, pages 1–27. Springer, Heidelberg, February 2014.



- [FV03] Pierre-Alain Fouque and Frédéric Valette. The doubling attack - why upwards is better than downwards. In Colin D. Walter, Çetin Kaya Koç, and Christof Paar, editors, *CHES 2003*, volume 2779 of *LNCS*, pages 269–280. Springer, Heidelberg, September 2003.
- [GJM<sup>+</sup>11] Raveen R. Goundar, Marc Joye, Atsuko Miyaji, Matthieu Rivain, and Alexandre Venelli. Scalar multiplication on Weierstraß elliptic curves from co-Z arithmetic. *Journal of Cryptographic Engineering*, 1(2):161–176, August 2011.
- [GJR18] Dahmun Goudarzi, Antoine Joux, and Matthieu Rivain. How to securely compute with noisy leakage in quasilinear complexity. In Thomas Peyrin and Steven Galbraith, editors, *ASIACRYPT 2018, Part II*, volume 11273 of *LNCS*, pages 547–574. Springer, Heidelberg, December 2018.
- [GRV16] Dahmun Goudarzi, Matthieu Rivain, and Damien Vergnaud. Lattice attacks against elliptic-curve signatures with blinded scalar multiplication. In Roberto Avanzi and Howard M. Heys, editors, *SAC 2016*, volume 10532 of *LNCS*, pages 120–139. Springer, Heidelberg, August 2016.
- [Ham20] Mike Hamburg. Faster montgomery and double-add ladders for short Weierstrass curves. *IACR TCHES*, 2020(4):189–208, 2020. <https://tches.iacr.org/index.php/TCHES/article/view/8681>.
- [HIM<sup>+</sup>13] Johann Heyszl, Andreas Ibing, Stefan Mangard, Fabrizio De Santis, and Georg Sigl. Clustering algorithms for non-profiled single-execution attacks on exponentiations. In Aurélien Francillon and Pankaj Rohatgi, editors, *CARDIS 2013*, volume 8419 of *LNCS*, pages 79–93. Springer, 2013.
- [HMA<sup>+</sup>08] Naofumi Homma, Atsushi Miyamoto, Takafumi Aoki, Akashi Satoh, and Adi Shamir. Collision-based power analysis of modular exponentiation using chosen-message pairs. In Elisabeth Oswald and Pankaj Rohatgi, editors, *CHES 2008*, volume 5154 of *LNCS*, pages 15–29. Springer, Heidelberg, August 2008.
- [HMH<sup>+</sup>12] Johann Heyszl, Stefan Mangard, Benedikt Heinz, Frederic Stumpf, and Georg Sigl. Localized electromagnetic analysis of cryptographic implementations. In Orr Dunkelman, editor, *CT-RSA 2012*, volume 7178 of *LNCS*, pages 231–244. Springer, Heidelberg, February / March 2012.
- [HMOV04] Darrel Hankerson, Alfred Menezes, and Scott Vanstone. *Guide to Elliptic Curve Cryptography*. January 2004.
- [HWCD08] Hüseyin Hisil, Kenneth Koon-Ho Wong, Gary Carter, and Ed Dawson. Twisted Edwards curves revisited. In Josef Pieprzyk, editor, *ASIACRYPT 2008*, volume 5350 of *LNCS*, pages 326–343. Springer, Heidelberg, December 2008.
- [IISO10] Masami Izumi, Jun Ikegami, Kazuo Sakiyama, and Kazuo Ohta. Improved countermeasure against address-bit DPA for ECC scalar multiplication. In Giovanni De Micheli, Bashir M. Al-Hashimi, Wolfgang Müller, and Enrico Macii, editors, *Design, Automation and Test in Europe, DATE 2010, Dresden, Germany, March 8-12, 2010*, pages 981–984. IEEE Computer Society, 2010.
- [IIT03a] Kouichi Itoh, Tetsuya Izu, and Masahiko Takenaka. Address-bit differential power analysis of cryptographic schemes OK-ECDH and OK-ECDSA. In Burton S. Kaliski Jr., Çetin Kaya Koç, and Christof Paar, editors, *CHES 2002*, volume 2523 of *LNCS*, pages 129–143. Springer, Heidelberg, August 2003.

- [IIT03b] Kouichi Itoh, Tetsuya Izu, and Masahiko Takenaka. A practical countermeasure against address-bit differential power analysis. In Colin D. Walter, Çetin Kaya Koç, and Christof Paar, editors, *CHES 2003*, volume 2779 of *LNCS*, pages 382–396. Springer, Heidelberg, September 2003.
- [ISW03] Yuval Ishai, Amit Sahai, and David Wagner. Private circuits: Securing hardware against probing attacks. In Dan Boneh, editor, *CRYPTO 2003*, volume 2729 of *LNCS*, pages 463–481. Springer, Heidelberg, August 2003.
- [IYTT03] Kouichi Itoh, Jun Yajima, Masahiko Takenaka, and Naoya Torii. DPA countermeasures by improving the window method. In Burton S. Kaliski Jr., Çetin Kaya Koç, and Christof Paar, editors, *CHES 2002*, volume 2523 of *LNCS*, pages 303–317. Springer, Heidelberg, August 2003.
- [Joy07] Marc Joye. Highly regular right-to-left algorithms for scalar multiplication. In Pascal Paillier and Ingrid Verbauwhede, editors, *CHES 2007*, volume 4727 of *LNCS*, pages 135–147. Springer, Heidelberg, September 2007.
- [Joy09] Marc Joye. Highly regular m-ary powering ladders. In Michael J. Jacobson Jr., Vincent Rijmen, and Reihaneh Safavi-Naini, editors, *SAC 2009*, volume 5867 of *LNCS*, pages 350–363. Springer, Heidelberg, August 2009.
- [JT09] Marc Joye and Michael Tunstall. Exponent recoding and regular exponentiation algorithms. In Bart Preneel, editor, *AFRICACRYPT 09*, volume 5580 of *LNCS*, pages 334–349. Springer, Heidelberg, June 2009.
- [KJJ99] Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In Michael J. Wiener, editor, *CRYPTO'99*, volume 1666 of *LNCS*, pages 388–397. Springer, Heidelberg, August 1999.
- [KPH<sup>+</sup>19] Jaehun Kim, Stjepan Picek, Annelie Heuser, Shivam Bhasin, and Alan Hanjalic. Make some noise: Unleashing the power of convolutional neural networks for profiled side-channel analysis. *IACR TCHES*, 2019(3):148–179, 2019. <https://tches.iacr.org/index.php/TCHES/article/view/8292>.
- [LHT16] Adam Langley, Mike Hamburg, and Sean Turner. Elliptic Curves for Security. RFC 7748, January 2016.
- [MBPK21] Naila Mukhtar, Lejla Batina, Stjepan Picek, and Yinan Kong. Fake it till you make it: Data augmentation using generative adversarial networks for all the crypto you need on small devices. Cryptology ePrint Archive, Report 2021/991, 2021. <https://eprint.iacr.org/2021/991>.
- [MFK<sup>+</sup>20] Naila Mukhtar, Apostolos P. Fournaris, Tariq Mahmood Khan, Charis Dimopoulos, and Yinan Kong. Improved hybrid approach for side-channel analysis using efficient convolutional neural network and dimensionality reduction. *IEEE Access*, 8:184298–184311, 2020.
- [MMMS22] Loïc Masure, Pierrick Méaux, Thorben Moos, and François-Xavier Standaert. Effective and efficient masking with low noise using small-mersenne-prime ciphers. Cryptology ePrint Archive, Paper 2022/863, 2022. <https://eprint.iacr.org/2022/863>.
- [Mon87] Peter L. Montgomery. Speeding the Pollard and elliptic curve methods of factorization. *Mathematics of Computation*, 48:243–264, 1987.
- [MOP07] Stefan Mangard, Elisabeth Oswald, and Thomas Popp. *Power analysis attacks - revealing the secrets of smart cards*. Springer, 2007.

- [MPF<sup>+</sup>22] Naila Mukhtar, Louiza Papachristodoulou, Apostolos P. Fournaris, Lejla Batina, and Yinan Kong. Machine-learning assisted side-channel attacks on RNS ECC implementations using hybrid feature engineering. In Josep Balasch and Colin O’Flynn, editors, *Constructive Side-Channel Analysis and Secure Design - 13th International Workshop, COSADE 2022, Leuven, Belgium, April 11-12, 2022, Proceedings*, volume 13211 of *Lecture Notes in Computer Science*, pages 3–28. Springer, 2022.
- [NC17] Erick Nascimento and Lukasz Chmielewski. Applying horizontal clustering side-channel attacks on embedded ECC implementations. In Thomas Eisenbarth and Yannick Teglia, editors, *Smart Card Research and Advanced Applications - 16th International Conference, CARDIS 2017, Lugano, Switzerland, November 13-15, 2017, Revised Selected Papers*, volume 10728 of *Lecture Notes in Computer Science*, pages 213–231. Springer, 2017.
- [NCOS16] Erick Nascimento, Lukasz Chmielewski, David Oswald, and Peter Schwabe. Attacking embedded ECC implementations through cmov side channels. In Roberto Avanzi and Howard M. Heys, editors, *SAC 2016*, volume 10532 of *LNCS*, pages 99–119. Springer, Heidelberg, August 2016.
- [NIS13] NIST. FIPS PUB 186-4: Digital Signature Standard (DSS), July 2013.
- [OMHT06] Elisabeth Oswald, Stefan Mangard, Christoph Herbst, and Stefan Tillich. Practical second-order DPA attacks for masked smart card implementations of block ciphers. In David Pointcheval, editor, *CT-RSA 2006*, volume 3860 of *LNCS*, pages 192–207. Springer, Heidelberg, February 2006.
- [Pag02] D. Page. Theoretical use of cache memory as a cryptanalytic side-channel. Cryptology ePrint Archive, Report 2002/169, 2002. <https://eprint.iacr.org/2002/169>.
- [PC15] Guilherme Perin and Lukasz Chmielewski. A semi-parametric approach for side-channel attacks on protected RSA implementations. In Naofumi Homma and Marcel Medwed, editors, *CARDIS 2015*, volume 9514 of *LNCS*, pages 34–53. Springer, 2015.
- [PCBP21] Guilherme Perin, Łukasz Chmielewski, Lejla Batina, and Stjepan Picek. Keep it unsupervised: Horizontal attacks meet deep learning. *IACR TCHES*, 2021(1):343–372, 2021. <https://tches.iacr.org/index.php/TCHES/article/view/8737>.
- [PGMP19] Thomas Prest, Dahmun Goudarzi, Ange Martinelli, and Alain Passelègue. Unifying leakage models on a Rényi day. In Alexandra Boldyreva and Daniele Micciancio, editors, *CRYPTO 2019, Part I*, volume 11692 of *LNCS*, pages 683–712. Springer, Heidelberg, August 2019.
- [PITM14] Guilherme Perin, Laurent Imbert, Lionel Torres, and Philippe Maurine. Attacking randomized exponentiations using unsupervised learning. In Emmanuel Prouff, editor, *COSADE 2014*, volume 8622 of *LNCS*, pages 144–160. Springer, Heidelberg, April 2014.
- [PR13] Emmanuel Prouff and Matthieu Rivain. Masking against side-channel attacks: A formal security proof. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 142–159. Springer, Heidelberg, May 2013.

- [PSB<sup>+</sup>18] Emmanuel Prouff, Remi Strullu, Ryad Benadjila, Eleonora Cagli, and Cecile Dumas. Study of deep learning techniques for side-channel analysis and introduction to ASCAD database. Cryptology ePrint Archive, Report 2018/053, 2018. <https://eprint.iacr.org/2018/053>.
- [RCB16] Joost Renes, Craig Costello, and Lejla Batina. Complete addition formulas for prime order elliptic curves. In Marc Fischlin and Jean-Sébastien Coron, editors, *EUROCRYPT 2016, Part I*, volume 9665 of *LNCS*, pages 403–428. Springer, Heidelberg, May 2016.
- [RIL19] Thomas Roche, Laurent Imbert, and Victor Lomné. Side-channel attacks on blinded scalar multiplications revisited. In Sonia Belaïd and Tim Güneysu, editors, *CARDIS 2019*, volume 11833 of *LNCS*, pages 95–108. Springer, 2019.
- [Riv11] Matthieu Rivain. Fast and regular algorithms for scalar multiplication over elliptic curves. Cryptology ePrint Archive, Report 2011/338, 2011. <https://eprint.iacr.org/2011/338>.
- [SI11] Werner Schindler and Kouichi Itoh. Exponent blinding does not always lift (partial) spa resistance to higher-level security. In Javier Lopez and Gene Tsudik, editors, *ACNS 11*, volume 6715 of *LNCS*, pages 73–90. Springer, Heidelberg, June 2011.
- [SWP03] Kai Schramm, Thomas J. Wollinger, and Christof Paar. A new class of collision attacks and its application to DES. In Thomas Johansson, editor, *FSE 2003*, volume 2887 of *LNCS*, pages 206–222. Springer, Heidelberg, February 2003.
- [TPP18] Michael Tunstall, Louiza Papachristodoulou, and Kostas Papagiannopoulos. Boolean exponent splitting. Cryptology ePrint Archive, Report 2018/1226, 2018. <https://eprint.iacr.org/2018/1226>.
- [WCPB20] Leo Weissbart, Lukasz Chmielewski, Stjepan Picek, and Lejla Batina. Systematic side-channel analysis of curve25519 with machine learning. *J. Hardw. Syst. Secur.*, 4(4):314–328, 2020.
- [WPB19] Leo Weissbart, Stjepan Picek, and Lejla Batina. One trace is all it takes: Machine learning-based side-channel attack on eddsa. In Shivam Bhasin, Avi Mendelson, and Mridul Nandi, editors, *Security, Privacy, and Applied Cryptography Engineering - 9th International Conference, SPACE 2019, Gandhinagar, India, December 3-7, 2019, Proceedings*, volume 11947 of *Lecture Notes in Computer Science*, pages 86–105. Springer, 2019.
- [WPP22] Lichao Wu, Guilherme Perin, and Stjepan Picek. The best of two worlds: Deep learning-assisted template attack. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2022(3):413–437, 2022.

## A Proofs of Lemmas 1 to 3

### A.1 Proof of Lemma 1

*Proof.* By definition of the  $\varepsilon$ -hiddenness property and since  $\Pi$  has a total number of  $N_\Pi$  operations, the leakage distribution output by  $\text{ExpReal}_1$  experiment is  $(N_\Pi \cdot \varepsilon)$ -close to the distribution output by the  $\text{ExpReal}_0$  experiment. The distribution output by the  $\text{ExpSim}_0$  is identical as the distribution output by the  $\text{ExpSim}_1$  experiment, which is  $\gamma_1$ -close to that of  $\text{ExpReal}_1$  and  $(\gamma_1 + N_\Pi \cdot \varepsilon)$ -close to that of  $\text{ExpReal}_0$ . We hence get (16).  $\square$

### A.2 Proof of Lemma 2

*Proof.* Lemma 2 holds from the standard reduction from noisy-leakage security to random-probing security put forward in [DDF14]. We specifically use the tighter variant of this result based on the relative error distance in the definition of noisy functions [PGMP19, Lemma 3]. According to this lemma, for any  $\delta$ -noisy function  $f$ , there exists a  $\delta$ -random probing function  $f'$  such that  $f(x) = g \circ f'(x)$  for every  $x$  and for some noisy function  $g$ . The noisy leakage functions  $f_i \in \mathcal{F}$  applied in  $\text{ExpReal}_1(\mathcal{A}, \Pi, \mathcal{F})$  are replaced by the corresponding random probing functions  $f'_i \in \mathcal{F}'$  in  $\text{ExpReal}_2(\mathcal{A}, \Pi, \mathcal{F}')$ . In order to construct the simulator  $\text{Sim}$ , we simply run  $\text{Sim}'(\Pi, \mathbf{X}, \mathcal{F}')$  which outputs a distribution which is  $\gamma_2$ -close to the random probing leakage output of  $\text{ExpReal}_2$ . Then  $\text{Sim}$  applies the  $g_i$  functions to this simulated leakage. Let us denote by  $(\ell_1^R, \dots, \ell_z^R)$  the output distribution of  $\text{ExpReal}_2$  and by  $(\ell_1^S, \dots, \ell_z^S)$  the output distribution of  $\text{ExpSim}_2$ , which satisfies

$$(\ell_1^R, \dots, \ell_z^R) \approx_{\gamma_2} (\ell_1^S, \dots, \ell_z^S). \quad (20)$$

According to the aforementioned lemma [PGMP19, Lemma 3], the output distribution of  $\text{ExpReal}_1$  is identically distributed to  $(g_1(\ell_1^R), \dots, g_z(\ell_z^R))$ . On the other hand, by construction of  $\text{Sim}$  the output of  $\text{ExpSim}_1$  is defined as  $(g_1(\ell_1^S), \dots, g_z(\ell_z^S))$ . Finally, (20) clearly implies

$$(g_1(\ell_1^R), \dots, g_z(\ell_z^R)) \approx_{\gamma_1} (g_1(\ell_1^S), \dots, g_z(\ell_z^S)) . \quad (21)$$

for some  $\gamma_1 \leq \gamma_2$  which concludes the proof.  $\square$

### A.3 Proof of Lemma 3

*Proof.* We demonstrate hereafter that there exists a simulator  $\text{Sim}$  such that for every input  $\mathbf{X}$ ,  $\mathbf{k} = (k_0, k_1, \dots, k_n)$ , and with  $\gamma_2$  satisfying (18), we have

$$\text{Sim}(\Pi, \mathbf{X}) \approx_{\gamma_2} \text{LeakageSampler}(\text{ComputTrace}'(\Pi, \mathbf{X}, \mathbf{k}), \mathcal{F}'). \quad (22)$$

Our demonstration proceeds in two steps: we first describe a simulator  $\text{Sim}^*$  that achieves a perfect simulation assuming that the leakage sampler reveals the entire computation trace of the program (*i.e.* the full inputs of all the operations) except for one iteration in the inner loop (from Step 4 to Step 10), *i.e.* for one value of the loop index  $j$ , where this non-leaking iteration might be different in each iteration of the outer loop, *i.e.* for each  $i$ . We then show that we can use  $\text{Sim}^*$  to construct our simulator  $\text{Sim}$  which outputs a perfect simulation of

$$\text{LeakageSampler}(\text{ComputTrace}'(\Pi, \mathbf{X}, \mathbf{k}), \mathcal{F}')$$

with probability  $\gamma_2$  and fails otherwise (which implies the lemma statement).

For each  $i$  between 0 and  $n - 1$ , let us denote by  $j_i^*$  the loop index value corresponding to the non-leaking inner loop iteration. Namely, for the  $i$ th iteration of the main loop, we assume that the operations in Steps 5, 7, 8 and 9 do not leak when  $j = j_i^*$ . All the other

operations, either inside the inner loop for  $j \neq j_i^*$ , or outside the inner loop, completely leak their input values. We explain hereafter how to complete a perfect simulation of such a leakage from  $\Pi$  (the program description) and  $\mathbf{X}$  the input tuple of algebraic variables. The corresponding simulator is denoted  $\text{Sim}^*(\Pi, \mathbf{X}, \mathcal{J})$  where  $\mathcal{J} = \{j_0^*, \dots, j_{n-1}^*\}$  is the set of non-leaking indexes.

All the operations in Step 1 (the `preprocess` procedure) have inputs and outputs derived from  $\mathbf{X}$  which makes all the leakage in this step perfectly simulable from  $\mathbf{X}$  (by simply evaluating `preprocess` on  $\mathbf{X}$ ). Then, for the first iteration  $i = 0$  of the main loop, the leakage of Step 3 can also be perfectly simulated from  $\mathbf{X}$ . The simulation  $\text{Sim}^*$  then generates  $k_0^j$  uniformly at random from  $[0, 2^w)$  for  $j \in [0, d] \setminus \{j_0^*\}$ . Steps 5, 7, 8 and 9 can thus be perfectly simulated for  $j \in [0, j_0^* - 1]$ . Note that for the randomization operations of the form  $X_{out} \leftarrow R(X_{in})$ , the computation trace output of the `ComputTrace'` sampler contains the pair of input output  $(X_{in}, X_{out})$  of the instruction, where the variable  $X_{out}$  is substituted for a fresh random value. This can be perfectly simulated by  $\text{Sim}^*$  by sampling fresh random values for all the outputs  $X_{out}$  (and keeping track of these sampled values for the forthcoming operations). For iteration  $j = j_0^*$ , no leakage must be simulated. For  $j = j_0^* + 1$ , all the input variables in this iteration are also output variables of a randomization operation  $R$ . They can therefore be perfectly simulated with uniform values. Following iterations  $j > j_0^* + 1$  can be perfectly simulated from these randomly sampled values and by picking  $k_0^j$  uniformly at random from  $[0, 2^w)$ .

*Remark 10.* We stress that one would not be able to additionally simulate the leakage in iteration  $j_0^*$  without knowledge of  $k_0$ . Indeed, by definition the share  $k_0^{j_0^*}$  satisfies

$$k_0^{j_0^*} = k_0 \oplus \bigoplus_{j \neq j_0^*} k_0^j .$$

Any leakage in the `indexes` procedure for  $j = j_0^*$  would (partially) depend on  $k_0^{j_0^*}$  and hence (together with the other  $k_0^j$ ) of  $k_0$ . In the same way, a leakage in the `swap` procedure would depend on the bit  $b_\ell^{j_0^*}$  which is derived from  $k_0^{j_0^*}$ . Furthermore, the leakage of any randomization operation in the loop iteration  $j_0^*$  would also depend on the bit  $b_\ell^{j_0^*}$  together with the leakage of other iterations. For instance, if the randomization of  $X_\ell^1$  would leak, then one would observe which random value is sampled as output of  $R(X_\ell^1)$  and consequently deduce whether a swap has occurred or not from the input variables of the next iteration (hence recovering  $b_\ell^{j_0^*}$ ).

For any following iteration  $i \geq 1$ , we consider two cases depending on whether  $j_{i-1}^* = d$  or not. Whenever  $j_{i-1}^* = d$ , the algebraic variables in input of iteration  $i$  are the output of randomization operations that do not leak. They can therefore be perfectly simulated with fresh random values independently of the previous leakage. Whenever  $j_{i-1}^* < d$ , the input algebraic variables of iteration  $i$  are taken from the simulation of randomization operation in the inner loop iteration  $j = d$  of the last main loop iteration  $i - 1$ . Now that we have explained how to perfectly simulate the input variables of iteration  $i$  we can proceed with a simulation of this iteration in the exact same way as for the first iteration  $i = 0$ . In particular,  $\text{Sim}^*$  generates  $k_i^j$  uniformly at random from  $[0, 2^w)$  for  $j \in [0, d] \setminus \{j_i^*\}$  and fresh random values for all the output of the randomization operations from which a perfect simulation is achieved as described above.

Our simulator  $\text{Sim}(\Pi, \mathbf{X})$  works as follows: it first draws some coin with probability  $\delta'$  for each operation in order to determine whether the operation leaks or not (*i.e.* whether the random probing function reveals its input or  $\perp$ ). For each iteration  $i$  of the main loop, we denote by  $\text{Fail}_i$  the event that at least one operation leaks in each iteration  $j$  of the inner loop. We further denote by  $\text{Fail}$  the union of the failure events for all the iterations, that is

$$\text{Fail} = \text{Fail}_0 \cup \text{Fail}_1 \cup \dots \cup \text{Fail}_{n-1} . \quad (23)$$

Whenever **Fail** occurs, the simulator **Sim** stops and the simulation fails. Whenever **Fail** does not occur, there exists at least one index  $j_i^*$  per iteration  $i$  for which the iteration  $j = j_i^*$  in the inner loop does not leak. The simulator **Sim** then runs  $\text{Sim}^*(\Pi, \mathbf{X}, \mathcal{J})$  with  $\mathcal{J} = \{j_0^*, \dots, j_{n-1}^*\}$  from which a perfect simulation is obtained for all the operations except the non-leaking iterations. From this perfect simulation **Sim** can then extract the values to be simulated according to the drawn random probing coins.

We thus get a simulator  $\text{Sim}(\Pi, \mathbf{X})$  which outputs a perfect distribution or fails with probability

$$\gamma_2 = \Pr(\text{Fail}) = 1 - \prod_{i=0}^{n-1} (1 - \Pr(\text{Fail}_i)) . \quad (24)$$

Let us denote by  $\text{Fail}_i^j$  the failure event that at least one operation leaks in the inner iteration  $j$  of the iteration  $i$  of the main loop. The failure event  $\text{Fail}_i$  occurs at iteration  $i$  if all the iterations of the inner loop leak at least for one operation (in Step among 5, 7, 8, and 9), that is

$$\Pr(\text{Fail}_i) = \Pr(\text{Fail}_i^0 \cap \text{Fail}_i^1 \cap \dots \cap \text{Fail}_i^d) = \prod_{j=0}^d \Pr(\text{Fail}_i^j) ,$$

where the product of probabilities holds from the mutual independence of the  $\text{Fail}_i^j$ 's. Each inner loop iteration is composed of  $m$  index operations in the procedure **indexes**,  $2m$  randomization operations, and  $2m$  copies in the **swap** procedure which makes a total of  $5m$  operations per iteration. For every pair  $(i, j)$ , the probability that none of these operations leak hence equals

$$1 - \Pr(\text{Fail}_i^j) = (1 - \delta)^{5m} .$$

From this we, get  $\Pr(\text{Fail}_i) = (1 - (1 - \delta)^{5m})^{d+1} \leq (5m\delta)^{d+1}$ , which finally yields

$$\Pr(\text{Fail}) = 1 - \prod_{i=0}^{n-1} (1 - \Pr(\text{Fail}_i)) = 1 - \left(1 - (1 - (1 - \delta)^{5m})^{d+1}\right)^n \leq n(5m\delta)^{d+1} . \quad (25)$$

□