The Mysteries of LRA: Roots and Progresses in Side-channel Applications

Jiangshan Long[1], Changhai Ou[1,*], Zhu Wang[2], Fan Zhang[3],

[1] School of Cyber Science and Engineering, Wuhan Univeristy, Hubei, China (longjiangshan@whu.edu.cn).

[2] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China.

[3] College of Computer Science and Technology, Zhejiang University, Hangzhou, China.

Abstract: Evaluation of cryptographic implementations with respect to side-channels has been mandated at high security levels nowadays. Typically, the evaluation involves four stages: detection, modeling, certification and secret recovery. In pursuit of specific goal at each stage, inherently different techniques used to be considered necessary. However, since the recent works of Eurocrypt2022 [1] [2] and Eurocrypt2024 [3], linear regression analysis (LRA) has uniquely become the technique that is well-applied throughout all the stages. In this paper, we concentrate on this silver bullet technique within the field of side-channel. First, we address the fundamental problems of why and how to use LRA. The discussion of nominal and binary nature explains its strong applicability. To sustain effective outcomes, we provide in-depth analyses about the design matrix, regarding the sample distribution of plaintext and the chosen polynomial degree. We summarize ideal conditions that totally avoid multicollinearity problem, and explore the novel evaluator-advantageous property of LRA by means of model diagnosis. Then, we trace the roots where we theoretically elaborate its connections with traditional side-channel techniques, including Correlation Power Analysis (CPA), Distance-of-Means analysis (DoM) and Partition Power Analysis (PPA), in terms of regression coefficients, regression model and coefficient of determination. Finally, we probe into the state-of-the-art combined LRA with the so-called collapse function, demonstrating its relationship with another refined technique, G-DoM. We argue that properly relaxing the definition of bit groups equally satisfies our conclusions. Experimental results are in line with the theory, confirming its correctness.

Keywords: side-channel evaluation; linear regresssion; CPA; DoM; collapsed function

1 Introduction

Cryptographers have traditionally designed cipher systems assuming that they would be realized in a closed environment which does not leak any information about the internal state. However, this long-standing security assumption is nullified when an adversary is enhanced with sophisticated side-channel analyses other than only black box access. Without strict evaluation, cryptosystem risks emitting informative signals such as power consumption [4] and electromagnetic radiation [5] to the external. These leakages statistically depend on sensitive intermediate variables of the running cryptographic algorithm and therefore provide a breach for undermining the protection.

In response to this new threat, security assessment schemes such as Common Criteria [6] and FIPS 140-3 [7] have mandated the evaluation of cryptographic implementations against side-channel analyses at high security levels. Early studies (e.g., [8][9][10]) have been carried out where it is found out that the

*Corresponding author.

illegal side-channel analysis also abides by the Shannon's communication model. As a result, the security evaluation is generally conducted in the following sequence:

(1) Leakage Detection. Leakage detection is primary in side-channel evaluation. Referring to Shannon's information theory, it answers the most fundamental question: "Is there an information leaking source?" A typical way for achieving this goal is resorting to the hypothesis testing in statistics. So far, popular examples include Welch's T-test [11], Pearson's ρ-test [12] and χ^2-test [13], among others. Searching time samples on the entire leakage trace, evaluators aim to identify all data-dependent leaking points, without concern for whether they can be exploited at a reasonable computational cost.

(2) Leakage Modeling. Confirming the existence of information source, the natural next step is to profile it in a detailed manner. Some may argue against the necessity of this stage, citing the non-profiled side-channel analyses. However, [14][15] find out that this kind of techniques fails to deal with injective targets, in terms of recall in classification theory. Under these circumstances, leakage modeling establishes behaviors of leakages and helps understand the features of a given implementation. It studies how secret data is modulated onto physical carrier signals such as electrical currents, and subsequently transmitted to the external. Practical tools for this stage relates to probability density estimation, including parameterized Gaussian template [16], semi-parameterized Gaussian mixture [17] and non-parameterized histogram [18].

(3) Leakage Certification. A notorious problem in modeling is that the estimation quality is heavily impacted by the capabilities of the modeler (e.g., the available number of leakage measurements and computational resources). Without a thorough examination, a given implementation may look "secure" in front of an inferior evaluator, but is in fact vulnerable to a more powerful adversary (i.e., a sense of false security). Besides, an ideal model should capture all components that contribute to the data-dependency. To address these issues, leakage certification is an essential post stage to the modeling by quantifying differences between estimated model and real leakages. It ensures a close approximation to the true leakage distribution, whilst guaranteeing unbiasedness. Practical tools for this purpose include distance-sampling based divergence [19] and moment-based statistical test [12].

(4) Secret Recovery. As the last stage of security evaluation, an hypothetical adversary (corresponding to the role of information destination in Shannon's information theory) tries to recover the modulated secret from leakages by the use of so-called side-channel distinguishers (corresponding to the role of demodulator). Most distinguishers base on the idea of maximum likelihood principal where the secret is the unknown parameter to be deduced. This final stage accomplishes the evaluation by applying concrete distinguishers (e.g., [20][21]) with well-established models on the detected leaking points.

At first glance, in pursuit of specific goal at each stage, inherently different techniques used to be considered. However, since the recent works in 2022 [1] [2] and 2024 [3], linear regression analysis (LRA) has uniquely become the technique that is well-applied throughout all the four stages! Besides this unexpected versatility, LRA additionally overcomes several challenges that are insurmountable for other techniques. A brief description will be provided in the next subsection.

1.1 Related Works

Here, we briefly introduce the advantages of LRA as demonstrated in practical applications of security evaluation.

(1) Explainable leakage detection. Despite extensive research over a long period, leakage detection still faces critical problems of low interpretability or high overhead. Non-specific techniques (e.g., [7]) allow black box assumption of low cost, but are incapable of answering whether the detected leakages are exploitable for side-channel distinguishers. In contrast, specific techniques (e.g., [6]) aim to demonstrate concrete attacks on identified leakages. Yet, the cost becomes prohibitive if the target is closed-source (as in the case with most commercial products), and the conclusions are not transferable across leaking points. In [3], the LRA-based leakage detection method introduces the concept of collapse function

and resolves both challenges. It provides interpretable results by precisely locating the leaking bits of intermediate variables.

(2) Effective leakage modeling. Originally suggested as a leakage modeling technique, [22] proves that under the ordinary least squares (OLS) loss function, LRA produces the minimal residuals whenever correct subkey is assumed. In terms of goodness-of-fit tests [23], it indicates that the model achieves optimality and accounts for the vast majority of variations in leakages. This modeling technique is regarded as best under Gaussian noise [24]. Later research of LRA [25] drives a breakthrough from gate-level to instruction-level leakage simulation. It enables prediction of leakages given a sequence of instructions, without the need to actually execute it.

(3) Completeness certification. In modern implementations, intermediate state can be large (e.g., 32-bit words are processed in parallel) and difficult to handle. Security evaluation based on an inaccurate leakage model will easily lead to flawed conclusions. Tackling this issue, [1] explores the F-test for explanatory variable selection. The authors introduce the novel concept of "model completeness", ensuring that all contributing factors of leakages are captured and modeled. Leveraging this, [2] successfully reverse-engineered the components in microarchitecture of commercial implementations.

(4) Robust distinguisher. Following the initial implications in [22], [26] officially adopts LRA as a side-channel distinguisher. Compared with others, this on-the-fly profiling attack additional gains an advantage of robustness (i.e., can be performed with few general assumptions about the leakages). Subsequent works [14] suggest to replace the regression strategy in LRA with a step-wise one to overcome several ineffectiveness. Yet, [25] argues that this improvement is sensitive to iterative order, prone to overfitting, and has greatly uncertainty of the finalised model.

1.2 Our Contributions

In view of the exceptional capabilities and potentials, this paper develops new insights into LRA within the field of side-channel. Our contributions are as follows:

1. We address the fundamental problems of why and how to use LRA. Our study of the nominal and binary nature elucidates its strong applicability in side-channel analysis. We provide in-depth discussions about the statistical characteristics of design matrix, and summarize ideal conditions that avoid severe consequences of multicollinearity. We explore a novel aspect of LRA – it is particularly advantageous for evaluators (i.e., an adversary of equal power cannot benefit the same) – by checking the residuals.
2. We explore the connections of LRA with traditional side-channel techniques including CPA, DoM and PPA. We theoretically prove that: (i) regression coefficients are equivalent to DoM's values, with the corresponding explanatory variables act as selection functions; (ii) coefficient of determination can be expressed in terms of CPA's correlation coefficients, adjusted by an extra factor - the *Proportion of Variations Explained* (PVE); (iii) regression model is based on the nature of PPA, which is the weighted sum of partitions of leakage measurements. Also, a further study on the Equal Images under different Subkeys (EIS) property of regression model is provided.
3. Finally, we probe into the state-of-the-art technique that is combined with LRA for enhanced capability - the collapse function. We first demonstrate the relationship of this combined method with another refined side-channel technique, the G-DoM. Then, we extend its definition. We argue that bit groups of random size, even with possible bit dropping, equally satisfy our theoretical conclusions.

1.3 Organization

The remaining of this paper is organized as follows: preliminaries including leakage model, CPA, PPA, DoM and LRA are introduced in Section 2. Nominality, design matrix and evaluator-advantageous prop-

erty are detailed in Section 3. We analyze connections of LRA with traditional side-channel techniques and recently emerged collapsed functions in Section 4. Experiments on PRESENT and AES are presented in Section 5. Finally, we conclude this paper in Section 6.

2 Preliminaries

2.1 Side-channel Leakages and Leakage Model

Side-channel analysis treats the secret as a tuple of subkeys. Let k^* denote the target subkey selected from a set $\mathcal{K}$ and k denote any possible guessing value. Let m denote the corresponding plaintext byte variable. Cryptographic algorithm keeps to group operation for closure property (e.g., all computations in AES take place on Galois field $\mathbb{F}_2^8$). Let $x = \mathcal{G}(m, k)$ denote one of such operations and x is the n-bit assumed intermediate variable. Let $\mathcal{F}$ denote the leakage function which describes the physical signals leaked according to the computation of $\mathcal{G}$, and whose distribution over m is identical for any subkey (i.e., the EIS property, see [22][27][28] for details). Let $\mathcal{L}$ denote the leakage measurement and

$$\mathcal{L} = \mathcal{F} \circ \mathcal{G}(m, k^*) + \mathsf{N} = \mathcal{F}(x^*) + \mathsf{N}, \tag{1}$$

where N is the independent Gaussian noise with $\mathbb{D}\{\mathsf{N}\} = \sigma_N^2$. Both the algebraic properties of the running cryptographic algorithm and the physical characteristics of the underlying hardware circuits determine the resistance of an implementation against side-channel attacks. This intuition is captured by $\mathcal{G}$ and $\mathcal{F}$ respectively and we do not assume any restrictions on them to make our results well-applied to any scenario.

2.2 Distance-of-Means Analysis

DoM targeting a single bit of x^* is the first proposed side-channel technique [20]. In this attack, leakage measurements are divided into two subsets according to the assumed value of the target bit (i.e., the selection function). Leakages bundled together are deemed to share the same distribution. Denoted as SB-DoM (single bit DoM), it is expressed as:

$$\begin{aligned} \mathbf{D}_{\mathsf{SB-DoM}} &= \arg\max_{k \in \mathcal{K}} \ \Delta^{(k)}_{\mathsf{SB-DoM}}(\mathrm{T}) \\ &= \arg\max_{k \in \mathcal{K}} \ \mathbb{E}\{\mathcal{L}_{q|x_q^{[\mathrm{T}]}=1}\} - \mathbb{E}\{\mathcal{L}_{q|x_q^{[\mathrm{T}]}=0}\}, \end{aligned} \tag{2}$$

where $\mathcal{L}_q$ is the measurement of the q-th encryption, x_q is the corresponding assumed intermediate value calculated under guessing subkey k and plaintext byte m_q, and $x_q^{[\mathrm{T}]}$ is the T-th bit (i.e., the target bit) of x_q. It is soon extended to multiple bits to overcome some algebraic property of cryptographic algorithm that lead to failure in the mono-bit setting [29]: the generalized DoM. Abbreviated as G-DoM, it is defined as:

$$\begin{aligned} \mathbf{D}_{\mathsf{G-DoM}} &= \arg\max_{k \in \mathcal{K}} \ \Delta^{(k)}_{\mathsf{G-DoM}} \\ &= \arg\max_{k \in \mathcal{K}} \ \mathbb{E}\{\mathcal{L}_{q|x_q < \lfloor \frac{n}{2} \rfloor}\} - \mathbb{E}\{\mathcal{L}_{q|x_q \geq \lceil \frac{n}{2} \rceil}\}. \end{aligned} \tag{3}$$

The idea of summing existing distinguishers to define a new one has been realized in [29] where SB-DoM on each bit of x is integrated. This attack is denoted as M-DoM (multiply bit DoM) and it is straightforwardly written as:

$$\begin{aligned} \mathbf{D}_{\mathsf{M-DoM}} &= \arg\max_{k \in \mathcal{K}} \ \Delta^{(k)}_{\mathsf{M-DoM}} \\ &= \arg\max_{k \in \mathcal{K}} \ \textstyle\sum_{i=1}^{n} \Delta^{(k)}_{\mathsf{SB-DoM}}(i). \end{aligned} \tag{4}$$

2.3 Correlation Power Analysis and Partition Power Analysis

CPA identifies subkey by assessing the linear fitting rate between leakage model and measurements. It implicitly extends the binary classification of DoM to a multiple one by incorporating the well-known Pearson's correlation coefficient. Generally, CPA can be conducted either in a profiled way [30], or in a non-profiled way with an a-priori leakage model (e.g., Hamming Weight) [21]. In this paper, we concentrate on the more powerful version of profiled CPA which is expressed as:

$$\mathbf{D}_{\mathsf{CPA}} = \underset{k \in \mathcal{K}}{\arg\max}\ \rho\left(\mathcal{L}_{q=1,2,\ldots,Q}, \hat{\mathcal{F}}(x_q)\right), \tag{5}$$

where $\hat{\mathcal{F}}$ denotes the profiled leakage function with some estimation errors. This improvement of multi-classification is then explicitly formalized in [31] by introducing PPA:

$$\mathbf{D}_{\mathsf{PPA}} = \underset{k \in \mathcal{K}}{\arg\max}\ \sum\nolimits_i \alpha_i \times \mathbb{E}\{\mathcal{L}_q | \hat{\mathcal{F}}\left(x_q\right) = \Omega_i\}. \tag{6}$$

$R(\hat{\mathcal{F}}) = \{\Omega_1, \Omega_2, \ldots\}$ is the range of $\hat{\mathcal{F}}$. As the name implies, PPA calculates a weighted sum of leakage partitions. Real constants α_i-s are to be determined. Taking Hamming weight as example, the parameters are chosen as (see Equ (7) in [31]): $R(\hat{\mathcal{F}}) = \{0, 1, \cdots, n\}$ and $\alpha_i = \frac{C_n^i}{2^n} \times (i - \sum_{j=0}^{n} \frac{C_n^j}{2^n} \times j)$.

2.4 Linear Regression Analysis

In essential, LRA performs a key-dependent basis decomposition for the leakage function $\mathcal{F}$, which can be seen as a Q-dimensional vector $\vec{\mathcal{F}}(.) = \{\mathcal{F}(x_1), \mathcal{F}(x_2), \ldots, \mathcal{F}(x_Q)\} \in \mathrm{R}^{\mathrm{Q}}$ in a real scenario. Here R^{Q} is the usual Euclidean space. The design matrix $\mathbf{G}$ consists of a series of binary explanatory column vectors, each of which is made up of instance values of the corresponding explanatory variable. It is written as:

$$\begin{aligned}\mathbf{G}(k) &= \{\vec{1}, \vec{x}^{[1]}, \vec{x}^{[2]}, \ldots, \vec{x}^{[1,2]}, \vec{x}^{[1,3]}, \ldots\} \\ &= \begin{bmatrix} 1 & x_1^{[1]} & x_1^{[2]} & \ldots & x_1^{[1,2]} & x_1^{[1,3]} & \ldots \\ 1 & x_2^{[1]} & x_2^{[2]} & \ldots & x_2^{[1,2]} & x_2^{[1,3]} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_Q^{[1]} & x_Q^{[2]} & \ldots & x_Q^{[1,2]} & x_Q^{[1,3]} & \ldots \end{bmatrix},\end{aligned} \tag{7}$$

where $x_q^{[w_1,w_2,\ldots]} = \prod_j x_q^{[w_j]}$. The decomposition mounts to finding a set of real coefficients β-s for the approximation:

$$\begin{aligned}\vec{\mathcal{F}}(.) \approx \beta_{-1} + \sum\nolimits_i \beta_i\, \vec{x}^{[i]} + \sum\nolimits_{i_1 \neq i_2} \beta_{i_1,i_2}\, \vec{x}^{[i_1,i_2]} + \\ \ldots + \sum\nolimits_{i_1 \neq i_2 \neq \cdots \neq i_d} \beta_{i_1,i_2,\cdots,i_d}\, \vec{x}^{[i_1,i_2,\cdots,i_d]},\end{aligned} \tag{8}$$

where $d \leq n$ is the polynomial degree. If $d = 1$, it indicates the common Independent Bit Leakage (IBL) [32], which matches most of devices in practice [26]. Let $\vec{\mathcal{L}} = \{\mathcal{L}_1, \ldots, \mathcal{L}_Q\} = \vec{\mathcal{F}}(.) + \vec{\mathsf{N}} \in \mathrm{R}^{\mathrm{Q}}$ denote the measurement column vector (i.e., the response vector) and $\vec{\beta}(k) = (\beta_{-1}, \beta_1, \beta_2, \ldots)^{\mathrm{T}}$ denote the coefficient vector we look for. Then, the optimal approximation in Equ (8) (written as $\vec{\mathcal{F}}(.) \approx \mathbf{G}(k)\vec{\beta}(k)$ in matrix form) equals to minimizing the convex residual function:

$$\begin{aligned}&||\vec{R}_s||^2 = ||\vec{\mathcal{L}} - \vec{\mathcal{F}}(.)||^2 = ||\vec{\mathcal{L}} - \mathbf{G}(k)\vec{\beta}(k)||^2 \\ &= ||\vec{\mathcal{L}}||^2 + \vec{\beta}^{\mathrm{T}}(k)\left(\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k)\right)\vec{\beta}(k) - 2\vec{\beta}^{\mathrm{T}}(k)(\mathbf{G}^{\mathrm{T}}(k)\vec{\mathcal{L}}).\end{aligned} \tag{9}$$

Notation $||.||$ denotes the L2 norm. The above quadratic form of residuals is minimal when its Jacobian is zero [33]. Assuming that $\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k)$ is invertible (otherwise see Section 3.2), Equ (9) exclusively attains its minimum at $\vec{\beta}(k) = (\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k))^{-1}\mathbf{G}^{\mathrm{T}}(k)\vec{\mathcal{L}}$. From a vector perspective, it manifests that $\vec{\mathcal{L}}$ is projected into the subspace spanned by $\mathbf{G}(k)$, with an orthogonal residual vector of minimum magnitude.

As a detection, certification or profiling method, the task is finished by obtaining the regression model under $k = k^*$ and performing statistical significance tests on the regression coefficients. But, an

additional goodness-of-fit test is required when LRA is used as a side-channel distinguisher. The sum of squared residuals, the coefficient of determination (R^2) and the Akaike information criterion are three typical ways for this purpose, aimed at finding the best-fitting $\mathbf{G}(k)$. Previous works [26] have validated their equivalence by stripping off key-independent terms. So, we adopt the most popular R^2:

$$\begin{aligned}\mathbf{D}_{\mathsf{LRA}} &= \arg\max_{k\in\mathcal{K}} R^2(k) \\ &= \arg\max_{k\in\mathcal{K}} 1 - \frac{||\vec{\mathcal{L}} - \mathbf{G}(k)(\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k))^{-1}\mathbf{G}^{\mathrm{T}}(k)\vec{\mathcal{L}}||^2}{(\vec{\mathcal{L}} - \bar{\mathcal{L}})^{\mathrm{T}}(\vec{\mathcal{L}} - \bar{\mathcal{L}})}.\end{aligned} \tag{10}$$

$\bar{\mathcal{L}}$ is the sample mean of the measurement set $\{\mathcal{L}_1, \ldots, \mathcal{L}_Q\}$. In the following, k is omitted if not specified. e is used to generally represent any explanatory variable and $\mathbf{G} = \{\vec{e_1}, \vec{e_2}, \ldots\}$.

3 LRA: Why, How and Who

In this section, we discuss the strong applicability of LRA within the field of side-channel. We investigate the statistical characteristics of design matrix and generalize ideal conditions for practical application. We demonstrate why LRA is advantageous for evaluators, through model diagnosis [34][35].

3.1 Why to Use LRA: The Nominal and Binary Nature

To put it plainly, the nominality of a variable refers to the "label" meaning of its values [36]. That is, different values are not comparable in numerical terms, absence of ordinality, and do not maintain any proportional relations. They are simply labels of classes for distinction and are interchangeable with each other. To be specific, the target bit variable in SB-DoM is not nominal, as swapping "0" and "1" will immediately result in a negative differential result. Naturally, this also applies to M-DoM, G-DoM, CPA and PPA.

In contrast, the nominality holds true for the binary explanatory variables in LRA. Swapping "0" and "1" of a specific variable does reverse its regression coefficient, but does not affect the others as well as the residuals and R^2 [37]. Furthermore, the binary property benefits from the basic knowledge that two points can always define a line. It explains the reason why "linear" regression is enough for side-channel analysis, even for those advanced scenarios where more sophisticated techniques such as mutual information analysis [38] and Cramér–von Mises test [17] are usually recommended. In a nutshell, LRA captures non-linear dependencies by breaking them down into pieces of linear dependencies defined by the corresponding binary explanatory variables. The nominality strengthens this ability, especially in terms of robustness.

Theorem 1. *The strong applicability of LRA in side-channel stems from the nominality and binary of explanatory variables.*

3.2 How to Use LRA: The Ideal Conditions for Application

In practice, the efficiency of LRA lies on the proper setup of $\mathbf{G}$. Otherwise, the positive semidefinite term $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ in Equ.(9) is singular. First, we study the row constraints with property:

Property 1. *For any guessing subkey, the rank of design matrix is bounded by the sample distribution of plaintext byte variable during encryptions, i.e., rank*$(\mathbf{G}(k)) \leq \#M \leq Q$.

Notation $\#M$ is the number of de-duplicated plaintext byte instance values during Q encryptions. This property is easily verified as repeated plaintexts will yield the same assumed intermediate value and therefore identical rows in $\mathbf{G}$, having no effect on the matrix rank due to Gaussian elimination.

Then, we turn our analysis to the column constraints of $\mathbf{G}$. In practice, a polynomial degree d is selected. This is natural because one major advantage of LRA is its few requirement for a-priori knowledge

about leakages. Once selected, LRA consistently exploits the same degree for all guessing subkeys. In this case, the number of columns is controlled as $\sum_{p=0}^{d}\binom{n}{p} \leq 2^n$, which directly leads to another property:

Property 2. *For any guessing subkey, the rank of design matrix is bounded by the number of columns under the selected degree d, i.e.,* $rank(\mathbf{G}(k)) \leq \sum_{p=0}^{d}\binom{n}{p} \leq 2^n$.

Summarizing above properties, we have the lemma:

Lemma 1. *The rank of design matrix falls into two cases:*

1. $rank(\mathbf{G}(k)) \leq \#M < \sum_{p=0}^{d}\binom{n}{p}$. *This indicates that for all subkeys the design matrices are rank deficient.*
2. $rank(\mathbf{G}(k)) \leq \sum_{p=0}^{d}\binom{n}{p} \leq \#M$. *This indicates that the design matrix has an opportunity to become full rank.*

There are serious consequences according to the lemma.

Corollay 1. *If the design matrix suffers from a rank deficiency, the explanatory column vectors encounter a complete multicollinearity problem, where their number exceeds the actual dimension of the column space they span.*

This immediately gives rise to the non-invertibility of $\mathbf{G}^{\mathrm{T}}\mathbf{G}$. Generally, multicollinearity enlarges standard errors of regression coefficients, widening their confidence intervals and making the result unreliable. A slight variation in measurements can significantly alter the estimated results. Moreover, R^2 is falsely high, as LRA mistakenly interprets the random noise as correlated effects rather than capturing the true variations.

Unfortunately, addressing only the rank issue is insufficient to totally resolve the multicollinearity problem. This is because linearly independent column vectors do not necessarily have zero Pearson's correlation coefficients among them. With infinite measurements, design matrices become full rank for any subkey k and degree d. The binary representation of the intermediate variable x traverses from all zeros to all ones. In this case, first order explanatory variables have zero correlation coefficients with each other: $\mathbb{E}(x^{[i]} \cdot x^{[j]}) = \mathbb{E}(x^{[i]}) \cdot \mathbb{E}(x^{[j]})$. However, this no longer holds for higher order ones:

$$\begin{aligned}\mathbb{E}(x^{[w_{i1},w_{i2},\ldots]} \cdot x^{[w_{j1},w_{j2},\ldots]}) &= \prod_{w\in Z\setminus\widetilde{Z}} \mathbb{E}(x^{[w]}) \prod_{w\in\widetilde{Z}} \mathbb{E}((x^{[w]})^2) \\ &= (\frac{1}{2})^{|Z|} \leq \mathbb{E}(x^{[w_{i1},w_{i2},\ldots]}) \cdot \mathbb{E}(x^{[w_{j1},w_{j2},\ldots]}),\end{aligned} \tag{11}$$

where the set $Z = \{w_{i1}, w_{i2}, \ldots\} \cup \{w_{j1}, w_{j2}, \ldots\}$ and $\widetilde{Z} = \{w_{i1}, w_{i2}, \ldots\} \cap \{w_{j1}, w_{j2}, \ldots\}$. Non-zero correlation coefficients can develop into a new multicollinearity problem.

Corollary 2. *Though of full rank, the design matrix may still exhibit a severe multicollinearity problem because of the non-zero correlation coefficients among explanatory variables.*

There are some statistical tools to help quantify the extent of multicollinearity. In this paper we recommend Variance Inflation Factor (VIF) [34]. The VIF value for an explanatory vector $\vec{e} \in \mathbf{G}$ is assessed by conducting an auxiliary regression:

1. The tested explanatory vector $\vec{e}$ replaces the leakage measurement vector $\vec{\mathcal{L}}$ to become the new response vector.
2. The remaining $\mathbf{G}\backslash\{\vec{e}\}$ forms the new design matrix.
3. Auxiliary regression: $R^{'2} = \text{Regression}\,(\mathbf{G}\backslash\{\vec{e}\}, \vec{e})$.
4. Variance inflation factor: $\text{VIF} = 1/(1 - R^{'2})$.

If VIF > 10, it generally implies that the tested explanatory vector can almost be re-expressed as a linear combination of the others, and thus indicates a strong multicollinearity.

To totally avoid multicollinearity problem, we sum up following conditions for ideal applications of LRA in practice:

Theorem 2. *The conditions for ideal applications of LRA in practice include: (i) a balanced or sufficiently large leakage measurement set; (ii) a limited degree parameter to* $d = 1$.

"Balanced" means the number of measurements for each possible plaintext byte value is constant. Note that under the first ideal condition (degree is not restricted to $d = 1$), there are additional interesting properties regarding the design matrix:

Property 3. *Under balanced or infinite measurements, the ranks of design matrices are identical across guessing subkeys, i.e.,* $\forall k_1, k_2 \in \mathcal{K}$, $rank(\mathbf{G}(k_1)) = rank(\mathbf{G}(k_2))$.

This property derives from the closure property of the cryptographic operation $\mathcal{G}$ described in Subsection 2.1. Taking the most common case $\mathcal{G}(m, k) = m \oplus k$ as example, one can establish the following relationship between guessing subkeys:

$$\begin{aligned} \mathcal{G}(p, k_1) &= p \oplus k_1 = (p \oplus k_1 \oplus k_2) \oplus k_2 \\ &= p^{'} \oplus k_2 = \mathcal{G}(p^{'}, k_2), \end{aligned} \tag{12}$$

which states that the rows of $\mathbf{G}(k_1)$ corresponding to plaintext byte value p are in fact identical to those of $\mathbf{G}(k_2)$ corresponding to plaintext byte value $p^{'}$. Similarly, we get:

$$\begin{aligned} \mathcal{G}(p^{'}, k_1) &= p^{'} \oplus k_1 = (p \oplus k_1 \oplus k_2) \oplus k_1 \\ &= p \oplus k_2 = \mathcal{G}(p, k_2). \end{aligned} \tag{13}$$

From the two preceding equations, converting $\mathbf{G}(k_1)$ to $\mathbf{G}(k_2)$ only involves a series of row swaps which will not change the matrix rank. Likewise, these elementary operations won't change the linear relationships among column vectors either:

Property 4. *Under balanced or infinite measurements, VIF-s among column vectors are identical across guessing subkeys.*

In a word, the first condition on measurements ensures full ranks of design matrices, while the second one on degree thoroughly eliminates multicollinearity. Some might consider the second one as less crucial than the first, since it appears to be merely a post improvement. However, we emphasize that the second condition itself holds a unique position of preventing overfitting and incapability of LRA. As d increases, the growing complexity of regression model enhances its ability to interpret the variations in leakages, gradually giving rise to an overfitting rather than genuine predictive progress. Supporting this, let us consider the worst case $d = n$ with the simplest balanced measurement set $Q = 2^n$ (i.e., only one measurement for each plaintext byte value). In this case, $\mathbf{G}(k)$-s of all subkeys turn into invertible square matrices and:

$$\forall k \in \mathcal{K},\ \mathbf{G}(k) \left(\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k)\right)^{-1} \mathbf{G}^{\mathrm{T}}(k) = E. \tag{14}$$

This equation suggests that LRA experiences serious overfitting, leaving no residuals and achieving a perfect fit with $R^2 = 1$. Moreover, LRA loses its ability to distinguish subkeys. Estimated outcomes belonging to different subkeys are identical, rendering them invalid and worthless. Doubling the number of measurements for each plaintext value by a constant factor does not affect the column space of design matrices. It will only gently decrease the fitting rate by introducing additional uncertainties, but still maintains the incapability:

Corollary 3. *Under balanced or infinite measurements, yet at an inappropriately high degree d, the estimated outcomes – including the coefficient of determination, regression coefficients and regression model – tend to consistency across subkeys, due to overfitting and incapability of LRA.*

3.3 Who Should Use LRA: Advantageous for Evaluators

The requirement of balanced or large measurement set has shown some signs of this technique of being advantageous for evaluators. Nevertheless, there are more evidences reflected on the statistical characteristics of residuals. Before illustrating, we emphasize that following reasoning is not confined to the ideal conditions. As starting point, consider the lemma:

Lemma 2. *The uniquely highest value of R^2 under the correct guessing subkey k^* is supported by the equation:*

$$\underset{\hat{\mathcal{F}}(.):\mathbb{F}_2^n \to \mathrm{R}}{\arg\min} \quad \mathbb{E}\{||\vec{\mathcal{L}} - \vec{\hat{\mathcal{F}}}(.)||^2\} = \vec{\mathcal{F}}(.). \tag{15}$$

The minimum value $\mathbb{E}\{||\vec{\mathcal{L}} - \vec{\hat{\mathcal{F}}}(.)||^2\} = \mathbb{E}\{||\vec{\mathcal{F}}(.) - \vec{\hat{\mathcal{F}}}(.)||^2\} + \mathbb{E}\{\mathsf{N}^2\} \geq \mathbb{E}\{\mathsf{N}^2\}$ is exclusively attained at $\mathcal{F} = \hat{\mathcal{F}}$.

This lemma is proved in [22]. Suppose an adversary who has any knowledge needed about the device, except for k^*. We demonstrate that even in the presence of such a powerful adversary, his benefits from LRA are limited. The residual in the q-th encryption consists of two parts: $\delta(x_q^*) = \mathcal{F}(x_q^*) - \hat{\mathcal{F}}(x_q) = (\mathcal{F}(x_q^*) - \hat{\mathcal{F}}(x_q^*)) + (\hat{\mathcal{F}}(x_q^*) - \hat{\mathcal{F}}(x_q)) = \delta_{\mathcal{F}}(x_q^*) + \delta_k(x_q^*)$. The first part captures estimation errors from random noise and can be arbitrarily small by supplying additional measurements. They are normal, homoscedastic and independent. The second part represents informative key-dependent biases. If $x_q \neq x_q^*$, leakages generated according to x_q^* won't be correctly modeled, and thus $\delta_k(x_q^*) \neq 0$. So,

Lemma 3. *For incorrect guessing subkeys $k \in \mathcal{K}\backslash\{k^*\}$, residuals $R_s = \{\delta(x_1^*), \delta(x_2^*), \ldots, \delta(x_Q^*)\}$ of LRA do not satisfy normality, homoscedasticity and independence property.*

This lemma comes directly from the non-random key-dependent offsets $\delta_k(x_q^*)$, added to the traditional normal, homoscedastic and independent residuals $\delta_{\mathcal{F}}(x_q^*)$. For the adversary, the strength of LRA is largely compromised. The distorted residuals may invalidate the hypothesis tests, such as T-tests (recommended for secret recovery in [14]) and F-tests (recommended for leakage detection in [3], leakage modeling in [25] and leakage certification in [1][2]). Without knowledge of k^* and access to balanced or sufficiently large measurement set, adversary cannot fully leverage the capabilities of LRA.

Theorem 3. *LRA is evaluator-advantageous because of requirements of sufficient measurements and knowledge of k^*.*

4 LRA: Roots and Progress

In this section, we explore the connections of LRA with traditional side-channel techniques under the ideal conditions of balanced measurements and $d = 1$. Our theoretical proofs elucidate the relationships between the regression coefficients and both M-DoM and SB-DoM, between the coefficient of determination and CPA, and between the regression model and PPA. We probe into the combined LRA with the state-of-the-art collapse functions, which leads us to revisit another refined technique, G-DoM. We extend the definition of collapse function and summarize the criteria it should follow.

4.1 Regression Coefficients and DoM

Recalling Equ (2) of SB-DoM, we deduce:

$$\mathbb{E}\{\mathcal{L}_{q|x_q^{[\mathrm{T}]}=1}\} = \frac{\mathbb{E}\{\mathcal{L}_{q|x_q^{[\mathrm{T}]}=1}\} \times \mathbb{P}\{x_q^{[\mathrm{T}]} = 1\}}{\mathbb{P}\{x_q^{[\mathrm{T}]} = 1\}} = 2\mathbb{E}\{\mathcal{L}_q x_q^{[\mathrm{T}]}\} \tag{16}$$

with $\mathbb{P}\{x_q^{[\mathrm{T}]} = 1\} = 1/2$. Similarly, we can obtain:

$$\mathbb{E}\{\mathcal{L}_{q|x_q^{[\mathrm{T}]}=0}\} = 2\mathbb{E}\{\mathcal{L}_q(1 - x_q^{[\mathrm{T}]})\}. \tag{17}$$

Hence, the differential value of SB-DoM is re-expressed as:

$$\Delta^{(k)}_{\mathsf{SB-DoM}}(\mathrm{T}) = 4\mathbb{E}\{\mathcal{L}_q x_q^{[\mathrm{T}]}\} - 2\mathbb{E}\{\mathcal{L}_q\}. \tag{18}$$

Then, we the turn analysis to the regression coefficients:

$$\vec{\beta}(k) = \{(\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k))^{-1}\} * \{\mathbf{G}^{\mathrm{T}}(k)\vec{\mathcal{L}}\}. \tag{19}$$

The second term of the equation is easily expanded as:

$$\begin{aligned}\mathbf{G}^{\mathrm{T}}(k)\vec{\mathcal{L}} &= \{\vec{\mathcal{L}}\cdot\vec{1},\ \ \vec{\mathcal{L}}\cdot\vec{x}^{[1]},\ \ \ldots,\ \ \vec{\mathcal{L}}\cdot\vec{x}^{[n]}\}^{\mathrm{T}} \\ &= \{\mathbb{E}\{\mathcal{L}_q\},\ \mathbb{E}\{\mathcal{L}_q x_q^{[1]}\},\ \cdots,\ \mathbb{E}\{\mathcal{L}_q x_q^{[n]}\}\}^{\mathrm{T}} \times Q.\end{aligned} \tag{20}$$

Comparing Equ (20) with (18), it appears that the differential values of SB-DoM can be described as linear weighted sums of elements from the second term vector of regression coefficients. Hence, the natural next step is to derive possible weights from the first term $(\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k))^{-1}$. We apply Gauss-Jordan elimination to examine the inverse matrix, appending an identity matrix to the right side of $\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k)$ and performing a series of elementary row operations. This transforms it into the identity matrix meanwhile converts the appended one into $(\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k))^{-1}$. The original augmented matrix is:

$$\left[\begin{array}{ccccc:ccccc} 1 & 1/2 & 1/2 & \ldots & 1/2 & 1 & 0 & 0 & \ldots & 0 \\ 1/2 & 1/2 & 1/4 & \ldots & 1/4 & 0 & 1 & 0 & \ldots & 0 \\ 1/2 & 1/4 & 1/2 & \ldots & 1/4 & 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1/2 & 1/4 & 1/4 & \ldots & 1/2 & 0 & 0 & 0 & \ldots & 1 \end{array}\right] \tag{21}$$

Note that we extract the factor Q^{-1} from the inverse matrix to compensate the term Q in Equ (20). The general rules of $\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k)$ are based on the distributions of binary variable:

1. The first row and column are produced by the scalar products with the constant vector $\vec{1}$. The only 1 results from the scalar product with itself, while 1/2-s result from the scalar product with other explanatory vectors.
2. The remaining submatrix: values on the principal diagonal equal to 1/2, arising from the self product of each explanatory vector; the other values are 1/4, corresponding to the results between different explanatory vectors.

We use r_i for the i-th row and r_{ij} for the j-th element of that row. We perform pivoting on the first row as the initial step:

1. $r_1 = (n+1) \times r_1 - 2 \times \sum_{i>1} r_i$,
2. $\forall i > 1,\ r_i = r_i - 1/2 \times r_1$.

It leads to the following intermediate matrix:

$$\left[\begin{array}{ccccc:ccccc} 1 & 0 & 0 & \ldots & 0 & n+1 & -2 & -2 & \ldots & -2 \\ 0 & 1/2 & 1/4 & \ldots & 1/4 & -\frac{n+1}{2} & 2 & 1 & \ldots & 1 \\ 0 & 1/4 & 1/2 & \ldots & 1/4 & -\frac{n+1}{2} & 1 & 2 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 1/4 & 1/4 & \ldots & 1/2 & -\frac{n+1}{2} & 1 & 1 & \ldots & 2 \end{array}\right]. \tag{22}$$

Then, the remaining steps of Gauss-Jordan elimination are loops of two substeps as follows ($2 \le i \le n$):

1. $r_i = 4/(n+3-i) \times \{(n+2-i) \times r_i - \sum_{j>i} r_j\}$,
2. $\forall j > i,\ r_j = r_j - 1/4 \times r_i$.

Let us focus on the appended matrix which will eventually turn into the inverse. Elements r_{ij} after the first substep adhere to:

1. **case 1**: $j = 1$. The element is calculated as: $r_{i1} = 4/(n+3-i) \times \{(n+2-i) \times (-\frac{n+3-i}{2}) - (n+2-i) \times (-\frac{n+3-i}{2})\} = -2$.
2. **case 2**: $2 \leq j \leq i-1$. The elements are calculated as: $r_{ij} = 4/(n+3-i) \times \{(n+2-i) \times 0 - (n+2-i) \times 0\} = 0$
3. **case 3**: $j = i$. The element is calculated as: $r_{ii} = 4/(n+3-i) \times \{(n+2-i) \times 2 - (n+2-i) \times 1\} = 4$
4. **case 4**: $i+1 \leq j \leq n$. The elements are calculated as: $r_{ij} = 4/(n+3-i) \times \{(n+2-i) \times 1 - (n+2-i) \times 1 - 1\} = 0$

After the second substep, elements r_{jp} $(j > i)$ adhere to:

1. **case 1**: $p = 1$. The element is calculated as:
   $r_{jp} = -(n+3-i)/2 - 1/4 \times (-2) = -(n+3-(i+1))/2$
2. **case 2**: $2 \leq p \leq i-1$. The elements are calculated as:
   $r_{jp} = 0 - 1/4 \times 0 = 0$
3. **case 3**: $p = i$. The element is calculated as:
   $r_{jp} = 1 - 1/4 \times 4 = 0$
4. **case 4**: $p = i+1$. The element is calculated as:
   $r_{jp} = 2 - 1/4 \times 0 = 2$
5. **case 5**: $i+2 \leq p \leq n$. The elements are calculated as:
   $r_{jp} = 1 - 1/4 \times 0 = 1$

Combining the above, the intermediate state looks like:

$$\begin{bmatrix} n+1 & -2 & -2 & \dots & -2 & \dots & -2 & \dots & -2 \\ -2 & 4 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ -2 & 0 & 4 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -2 & 0 & 0 & \dots & 4 & \dots & 0 & \dots & 0 \\ -\frac{n+3-(i+1)}{2} & 0 & 0 & \dots & 0 & 2 & 1 & \dots & 1 \\ -\frac{n+3-(i+1)}{2} & 0 & 0 & \dots & 0 & 1 & 2 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\frac{n+3-(i+1)}{2} & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 2 \end{bmatrix}. \tag{23}$$

Comparing (22) and (23), both matrices share the similar submatrix in the lower right corner. Consequently, the final inverse matrix can be deduced to take the form:

$$(\mathbf{G}^{\mathrm{T}}(k)\mathbf{G}(k))^{-1} = \begin{bmatrix} n+1 & -2 & -2 & \dots & -2 \\ -2 & 4 & 0 & \dots & 0 \\ -2 & 0 & 4 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -2 & 0 & 0 & \dots & 4 \end{bmatrix}. \tag{24}$$

With this conclusion, we are able find out the relationships between the regression coefficients and the differential values:

$$\begin{aligned} \beta_{-1}(k) &= \mathbb{E}\{\mathcal{L}_q\} - \frac{1}{2}\Delta^{(k)}_{\mathsf{M-DoM}}, \\ \beta_i(k) &= \Delta^{(k)}_{\mathsf{SB-DoM}}(i),\ 1 \leq i \leq n. \end{aligned} \tag{25}$$

This equation under balanced measurements establishes the theorem for more practical scenarios of random measurements:

Theorem 4. *Under $d = 1$ and random measurements, the regression coefficients of LRA asymptotically converge to the differential values of SB-DoM, where the corresponding explanatory variables serve as the selection functions. The intercept converges to the difference between sample mean of measurements and half the differential value of M-DoM.*

This theorem provides a direct answer to the experimental phenomena observed in [14], explaining why the regression coefficients fluctuate more wildly under incorrect subkeys.

4.2 Regression Model and PPA

Based on **Theorem 4** and Equ (25), the regression model of LRA is now written and simplified as:

$$\begin{aligned}\hat{\mathcal{F}}(\mu) &= \mathbb{E}\{\mathcal{L}_q\} - \frac{1}{2}\Delta^{(k)}_{\mathsf{M-DoM}} + \textstyle\sum_{i=1}^{n} \Delta^{(k)}_{\mathsf{SB-DoM}}(i)\,\mu^{[i]} \\ &= \mathbb{E}\{\mathcal{L}_q\} + \textstyle\sum_{i=1}^{n} \Delta^{(k)}_{\mathsf{SB-DoM}}(i)(\mu^{[i]} - \frac{1}{2}) \\ &= \mathbb{E}\{\mathcal{L}_q\} + \frac{2}{Q}\textstyle\sum_{i=1}^{n}\sum_{p=1}^{Q}(\mu^{[i]} - \frac{1}{2})(\mathcal{L}_{p|x_p^{[i]}=1} - \mathcal{L}_{p|x_p^{[i]}=0}).\end{aligned} \tag{26}$$

The observations about the above equation are:

1. Given a certain p, the term $\mathcal{L}_{p|x_p^{[i]}=1} - \mathcal{L}_{p|x_p^{[i]}=0}$ can only yield two possible outcomes: $\mathcal{L}_p$ or $-\mathcal{L}_p$.
2. μ is the intermediate value being modeled. The term $\mu^{[i]} - \frac{1}{2}$ can also yield two possible outcomes: $1/2$ or $-1/2$.

Therefore, the contribution of a certain leakage measurement $\mathcal{L}_p$ to the above summation falls in two situations:

1. $\mu^{[i]} \oplus x_p^{[i]} = 0$. The $\mathcal{L}_p$ contributes positively.
2. $\mu^{[i]} \oplus x_p^{[i]} = 1$. The $\mathcal{L}_p$ contributes negatively.

Given this analysis, Equ (26) is further simplified as:

$$\begin{aligned}\hat{\mathcal{F}}(\mu) &= \mathbb{E}\{\mathcal{L}_q\} + \frac{1}{Q}\textstyle\sum_{p=1}^{Q}\{\sum_{i=1}^{n}(1 - x_p^{[i]} \oplus \mu^{[i]}) - x_p^{[i]} \oplus \mu^{[i]}\}\mathcal{L}_p \\ &= \frac{1}{Q}\textstyle\sum_{p=1}^{Q}(1 + n - 2 \times \sum_{i=1}^{n} \mu^{[i]} \oplus x_p^{[i]}) \times \mathcal{L}_p \\ &= \frac{1}{2^n}\textstyle\sum_{a\in\mathbb{F}_2^n}(1 + n - 2 \times \mathrm{HW}(a \oplus \mu)) \times \mathbb{E}\{\mathcal{L}_{p|x_p=a}\} \\ &= \frac{1}{2^n}\textstyle\sum_{a\in\mathbb{F}_2^n} c(a) \times \mathbb{E}\{\mathcal{L}_{p|x_p=a}\}.\end{aligned} \tag{27}$$

It is now clear that the regression model value is a weighted combination of measurement partitions. Specifically, LRA first divides leakages into 2^n partitions based on values of the intermediate variable. Then, the weight of partition is calculated as: $c(a) = 1 + n - 2 \times \mathrm{HW}(a \oplus \mu)$, depending on both the value currently being modeled (i.e., μ) and the key-dependent label (i.e., a) corresponding to the partition. The only difference between the regression models under different subkeys lies in the weight $c(a)$. Such type of computation fundamentally equal to the core essence of the earlier technique PPA.

Theorem 5. *Under $d = 1$ and random measurements, the regression model of LRA asymptotically converges to the form of PPA – a linear combination of leakage measurement partitions characterized by key-dependent weighted coefficients.*

This facilitates an in-depth discussion about the EIS property of the regression model. As stated in subsection 2.1, the property indicates that the output distribution of leakage function remains consistent

across subkeys. Specifically, they only decide the input-to-output mapping without affecting the image. The first moment of the regression model satisfies:

$$\begin{aligned}
&\mathbb{E}\{\hat{\mathcal{F}}(.)\} = \frac{1}{2^n}\sum_{\mu\in\mathbb{F}_2^n}\hat{\mathcal{F}}(\mu) \\
&= (\frac{1}{2^n})^2\sum_{a\in\mathbb{F}_2^n}[\sum_{\mu\in\mathbb{F}_2^n}(1+n-2\times \mathrm{HW}(a\oplus\mu))]\times\mathbb{E}\{\mathcal{L}_{p|x_p=a}\} \\
&= \frac{1}{2^n}\sum_{a\in\mathbb{F}_2^n} cst\times\mathbb{E}\{\mathcal{L}_{p|x_p=a}\} = \mathbb{E}\{\mathcal{L}_p\}.
\end{aligned} \tag{28}$$

The constant term $cst = (2^n)^{-1}\sum_{\mu\in\mathbb{F}_2^n}(1+n-2\times\mathrm{HW}(a\oplus\mu)) = 1$. For any given subkey, the first moment of regression model always equals to the sample mean of leakage measurements. However, it can be verified that this conclusion no longer stands for higher order moments. Thus, we have:

Corollary 4. *The regression model of LRA does not conform to the common EIS property of leakage model.*

This drawback in model profiling will expose in the following subsection where LRA is used a distinguisher.

4.3 Coefficient of Determination and CPA

As introduced, CPA identifies subkeys by utilizing the Pearson's linear correlation coefficient, while LRA employs the coefficient of determination. Although both are linear, their relationships within the field of side-channel remain ambiguous for many years. Now, we explicitly address this long standing issue, by expressing one in terms of the other:

$$\begin{aligned}
R^2 &= 1 - \frac{\sum_{q=1}^{Q}(\mathcal{L}_q - \hat{\mathcal{F}}(x_q))^2}{\sigma^2(\mathcal{L}_q)} \\
&= 2\times\frac{\sigma(\hat{\mathcal{F}}(x_q))}{\sigma(\mathcal{L}_q)}\times\frac{\mathbb{E}\{\mathcal{L}_q\hat{\mathcal{F}}(x_q)\} - \mathbb{E}\{\mathcal{L}_q\}\mathbb{E}\{\hat{\mathcal{F}}(x_q)\}}{\sigma(\mathcal{L}_q)\sigma(\hat{\mathcal{F}}(x_q))} \\
&\quad - \frac{\sigma^2(\hat{\mathcal{F}}(x_q)) + (\mathbb{E}\{\hat{\mathcal{F}}(x_q)\} - \mathbb{E}\{\mathcal{L}_q\})^2}{\sigma^2(\mathcal{L}_q)} \\
&= \frac{\sigma(\hat{\mathcal{F}}(x_q))}{\sigma(\mathcal{L}_q)}(2\rho(\mathcal{L}_q,\hat{\mathcal{F}}(x_q)) - \frac{\sigma(\hat{\mathcal{F}}(x_q))}{\sigma(\mathcal{L}_q)}).
\end{aligned} \tag{29}$$

Note that $(\mathbb{E}\{\hat{\mathcal{F}}(x_q)\}-\mathbb{E}\{\mathcal{L}_q\})^2 = 0$ due to Equ (28). As shown, R^2 is now defined as the linear correlation coefficient between the regression model and leakage measurements. The primary difference derives from the extra term $\sigma(\hat{\mathcal{F}}(x_q))/\sigma(\mathcal{L}_q)$, standing for the *Proportion of Variations Explained* (PVE). Compared to the total variations in leakages, this goodness-of-fit metric quantifies the part modeled by the regression model. Recalling **Corollary 3** and **4**, the second moment (i.e., variance) of $\hat{\mathcal{F}}$ varies across different subkeys, making the PVE term indispensable in the equation. Meanwhile, potential overfitting of LRA may result in misleading PVE values. These explain the experimental phenomena observed in [26] that LRA underperforms CPA as a distinguisher.

Theorem 6. *Under $d = 1$ and random measurements, the coefficient of determination in LRA asymptotically converges to CPA's correlation coefficients, adjusted by an extra factor - the Proportion of Variance Explained (PVE).*

4.4 Collapsed Function and LRA

With the advancement of parallel computing, a single leaking point can originate from several subkeys, which are hard to predict. This emerging challenge raises a new demand of scalability for side-channel techniques, especially in the applications of leakage detection and certification. If the number of leaking bits is affordable, LRA can be conducted on multiple subkeys without any modifications. However, the computation cost becomes prohibitive for highly parallel devices (i.e., too many coefficients in Equ (8)

need to be estimated). In this case, the collapse function introduces an elegant way for addressing this issue. Let ε denote the size of bit group to collapse, and it is usually a constant in the application. The collapse function on the input bit group $\{b_1, b_2, \ldots, b_\varepsilon\}$ is defined as:

$$Coll\left(\{b_1, b_2, \ldots, b_\varepsilon\}\right) = \begin{cases} 0, & \text{if } (b_1|b_2|\ldots|b_\varepsilon)_{(2)} < 2^{\varepsilon-1} \\ \\ 1, & \text{if } (b_1|b_2|\ldots|b_\varepsilon)_{(2)} \geq 2^{\varepsilon-1} \end{cases}. \tag{30}$$

Notation "|" denotes bit concatenation and subscript $(.)_{(2)}$ means the value of binary representation. In a word, function *Coll* maps a group of ε bits into a single binary explanatory variable, whilst retaining sufficient dependency. After this, LRA is directly carried out on the collapsed output which largely reduces the complexity from 2^ε to 1. As hinted by **Theorem 4**, the collapsed bit serves as the new selection function in the binary classification of DoM. Recalling Equ (3) and based on the analysis in [39], we derive the theorem:

Theorem 7. *Under $d = 1$ and random measurements, the estimated outcomes of LRA on collapsed bits, including the regression coefficients, regression model and coefficient of determination, equally satisfy **Theorem 4 $\sim$ 6**. The original LRA is a special case where each bit is collapsed to itself.*

This state-of-the-art technique can be formally expressed as:

$$\begin{aligned} \hat{\mathcal{F}}(x) = \mathbb{E}\{\mathcal{L}_q\} - \frac{1}{2}\textstyle\sum_i \Delta^{(k)}_{\mathsf{G-DoM}}(\{b_1, b_2, \ldots, b_\varepsilon\}_i) \\ + \textstyle\sum_i \Delta^{(k)}_{\mathsf{G-DoM}}(\{b_1, b_2, \ldots, b_\varepsilon\}_i) \times Coll\left(\{b_1, b_2, \ldots, b_\varepsilon\}_i\right), \end{aligned} \tag{31}$$

where the bit groups of the intermediate variable x satisfy:

$$\begin{aligned} \textstyle\bigcup_i \{b_1, b_2, \ldots, b_\varepsilon\}_i &= \{x^{[1]}, x^{[2]}, \ldots, x^{[n]}\}, \\ \textstyle\bigcap_i \{b_1, b_2, \ldots, b_\varepsilon\}_i &= \varnothing. \end{aligned} \tag{32}$$

The differential value of G-DoM is calculated as:

$$\begin{aligned} \Delta^{(k)}_{\mathsf{G-DoM}}(\{b_1, b_2, \ldots, b_\varepsilon\}_i) = \mathbb{E}\{\mathcal{L}_{q|Coll\left(\{b_1, b_2, \ldots, b_\varepsilon\}_{i,q}\right)=1}\} \\ - \mathbb{E}\{\mathcal{L}_{q|Coll\left(\{b_1, b_2, \ldots, b_\varepsilon\}_{i,q}\right)=0}\}, \end{aligned} \tag{33}$$

where $\{b_1, b_2, \ldots, b_\varepsilon\}_{i,q}$ represents the instance of the bit group $\{b_1, b_2, \ldots, b_\varepsilon\}_i$ in the q-th encryption. It is noteworthy that **Theorem 7** is independent of the parameter n, suggesting that it is well-applied to any cryptographic algorithm (e.g., $n = 8$ for AES S-box and $n = 4$ for PRESENT S-box) and arbitrary bit lengths of multiple intermediate variables (e.g., as initiated in [3] where x is the simple concatenation of several subkeys). Also, the theorem is unrelated to the parameter ε. This is because each bit group is processed independently in its corresponding G-DoM distinguisher and is transparent to each other. Therefore, the bit groups can be of any size, with inconsistent lengths, and may even undergo bit dropping. Furthermore, the collapse is set-oriented which is not confined to adjacent bits. These extensions of definition significantly boosts the flexibility in practical applications. At last, we generalize criteria that collapse function should adhere to:

Corollary 5. *The criteria for collapse function include: (i) mutually independent collapsed bits; (ii) uniformly distributed collapsed bits (the same occurrence probability of 0 and 1).*

5 Experimental Analyses

In this section, we provide detailed validation for our theory. Specifically, the experiments are carried out in three aspects:

1. Analyses of design matrices. This experiment centers around **Theorem 2** and **Corollary 1∼3**. We examine the effects of the measurement size Q, the degree d and the bit-length n on the matrices' ranks and multicollinearity.

2. Analyses of residuals. This experiment focuses on **Theorem 3**. We employ simulated leakages to ensure tight control over the explanatory variables. It is necessary because it rules out the possibility that experimental results could be induced by potential unknown explanatory variables.

3. Analysis of LRA's estimated outcomes (regression coefficients, regression model and coefficient of determination). This experiment focuses on **Theorem 4∼7**, and considers both the original LRA and the collapsed function enhanced LRA, using real leakages from AES and PRESENT.

5.1 Experimental Analyses on Design Matrices

Experimental results of the first experiment are displayed in Figs. 1∼9. We investigate both the AES ($n = 8$) and PRESENT ($n = 4$) S-boxes. Real leakages of AES are from the open dataset DPAcontest v4 [40]. Real leakages of PRESENT are sampled from an ATMega328p micro-controller whose clock frequency is 16 MHz. We apply a WaveRunner 8104 oscilloscope with a sampling rate of 1 GS/s. In the experiments, we define two metrics for illustration. The rank ratio (RT) is defined as the ratio of the current rank to the full rank, i.e., $\mathrm{RT}(\mathbf{G}) = rank(\mathbf{G})/\sum_{p=0}^{d}\binom{n}{p}$. It shows how the matrix rank grows with the number of measurements Q intuitively. To study statistical characteristics, 500 repeated experiments are conducted for each value of Q, and we provide the upper bound $\mathrm{RT}_{upper}(\mathbf{G})$, the lower bound $\mathrm{RT}_{lower}(\mathbf{G})$ and the average value $\mathrm{RT}_{avg}(\mathbf{G})$ simultaneously to ensure a comprehensive analysis. The model distance M^2 is defined as the squared difference between the regression model $\hat{\mathcal{F}}$ under k^* and that of $\hat{\mathcal{F}}'$ under an incorrect k, i.e., $M^2(k) = \sum_{m\in\mathbb{F}_2^n}(\hat{\mathcal{F}}\circ\mathcal{G}(m,k) - \hat{\mathcal{F}}'\circ\mathcal{G}(m,k))^2$. It quantifies the distances between regression models. Our observations are:

1. Growths of matrix ranks are more rapid for small n and d. With randomly encrypted plaintext bytes, AES (see Figs. 1(a)∼3(a)) under the settings of $d = 1$, 4 and 7 requires 18, 310 and 1270 leakage measurements respectively for the lower bounds to reach full ranks. Similar phenomena are found for PRESENT (see Figs 5(a)∼6(a)) where the number of measurements are 13 and 142 under $d = 1$ and 4. This observation is natural because large n and d lead to more columns in the design matrix, and thereby makes it more challenging to get rid of rank deficiency. It supports our **Theorem 2** that sufficient or balanced measurements are needed for LRA.

2. Distributions of matrix ranks are more dispersed for small n and d (see Figs. 1(a)∼3(a) and 5(a)∼6(a)). This observation is adequately explained by **Lemma 1**. On one hand, $rank(\mathbf{G})$ is bounded by the number of de-duplicated plaintext byte values $\#M$ during the encryptions. A small bit-length n brings a narrow value range. Consequently, for a certain encryption, the currently encrypted value is likely to collide with those previous ones, which contributes a redundant row to the design matrix, and thereby decreases the lower bound. On the other hand, increasing the degree d, due to the data dependency shown in **Corollary 2**, the newly added explanatory variables will equalize the influence on the matrix rank. Small variations of plaintext instance values in repeated experiments will no longer cause significant changes on the overall relationships among the column vectors, maintaining tighter bounds of matrix rank.

3. Multicollinearity problem is extensively encountered in the practical applications of LRA. In each repeated experiment, we calculate VIF for each explanatory vector $\vec{e} \in \mathbf{G}$, and the maximum value $\mathrm{VIF}_{\max}$, which represents the worst case, is displayed in Figs. 1(b)∼3(b) and 5(b)∼6(b). The number of measurements Q is set to the level at which the corresponding lower bound of rank ratio just achieves 1. Also, we allows for a slight excess of this level to avoid potential statistical biases. As shown, though VIF presents a negative correlation with Q, only when $d = 1$ does the $\mathrm{VIF}_{\max}$

fall below 10, regardless of AES or PRESENT. It directly confirms our **Corollary 2** that achieving full rank itself is far from a sufficient condition to eliminate the multicollinearity.

4. The VIF values rise with a higher degree d. For both AES and PRESENT, this observation underscores a careful selection of the degree parameter. From an ideal perspective, it is best to set d to 1 — under this setting, the probability of a multicollinearity problem occurring is only $5/500 = 1\%$ for AES (see Fig. 1(b)), and it further drops to 0% for PRESENT (see Fig. 5(b)). Besides highlighting the ideal conditions in **Theorem 2**, it also substantiates **Theorem 3** that an adversary with restricted leakage measurements cannot make full use of LRA.

5. Multicollinearity impairs statistical inference. Fig. 9 exhibits the confidence intervals for the regression coefficients under AES with $d = 1$, along with the VIF of each explanatory vector. In the left figure, $Q = 20$ is the minimum number of measurements required for a full rank design matrix, as shown in Fig 1(a). As depicted, the larger the VIF, the wider the confidence interval. Multicollinearity diminishes the precision of the estimated results, rendering them unstable and susceptible to minor changes. Also, it elevates the standard errors, undermining the reliability of statistical significance tests. The varying lengths of confidence intervals with centers all at zero make it difficult to identify which variables genuinely affect the leakages. They are disastrous for the application of LRA in leakage detection, modeling and certification. In the right figure, increasing Q to 120 has efficiently mitigated these problems: the confidence intervals are of almost equal length, and those far from zero indicate more significant contributions.

6. If d is inappropriate, LRA may face an overfitting or incapability even under balanced leakage measurements. In Figs. 4∼7, we set the balanced measurements to $Q = 256 \times num$ for AES and $Q = 16 \times num$ for PRESENT. We use box plot to exhibit the distributions of $R^2(k)$ and $M^2(k)$ over guessing subkey k. Apparently, a concentrated distribution of R^2-s or small values of M^2-s indicate similar behaviors of LRA across subkeys. As illustrated, R^2-s climb with the degree d, suggesting a possible overfitting. Supporting this, the declines of R^2-s in front of more complicated measurement sets demonstrate that the model fails to adequately capture the increased variations and uncertainties. Note that the phenomenon of $R^2 = 1$ under $num = 1$ agrees with our reasoning in Equ (14). Moreover, for AES, the distributions of R^2-s initially get dispersed and then turns compact. Similarly, the values of M^2-s first get rise and then falls. This is because simple models of low degree can only exhibit limited differences (resulting in compact R^2-s and small M^2-s). As the complexity increases, regression models under different subkeys begin to deviate from each other (producing dispersed R^2-s and high M^2-s). Yet, when d approaches its maximum, LRA converges to a state of incapability, making all R^2-s identical and M^2-s drop to zero (as seen in the single lines of box plot at $d = 8$). In contrast, the R^2-s and M^2-s present a monotone trend for PRESENT, due to its small maximum degree. This observation aligns with **Corollary 3** and advises the ideal conditions in **Theorem 2** to be both guaranteed.

7. High degrees d are generally undesirable even with infinite leakage measurements. While Figs. 1∼3 and 5∼6 illustrate the situations where design matrices just reach full rank, Fig. 8 relaxes Q to infinity to explore the suitability of large d for more powerful adversaries or evaluators. According to **Property 1**, this equates to a balanced measurement set. As shown, the results indicate that only $d = 1, 2$ are viable for AES (i.e., $\mathrm{VIF}_{\max} \leq 10$), whereas for PRESENT $d = 1, 2, 3$ are allowed. This highlights the general infeasibility of high degrees, even with infinite measurements, and emphasizes that both conditions in **Theorem 2** should be met simultaneously.

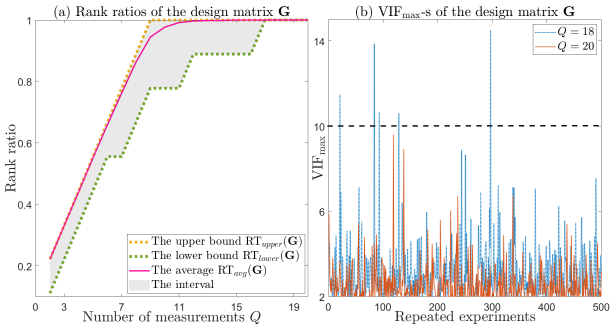

Figure 1: Statistical characteristics of the design matrix under AES ($n = 8$) and $d = 1$: (a): rank ratios; (b): VIF$_{\max}$-s.

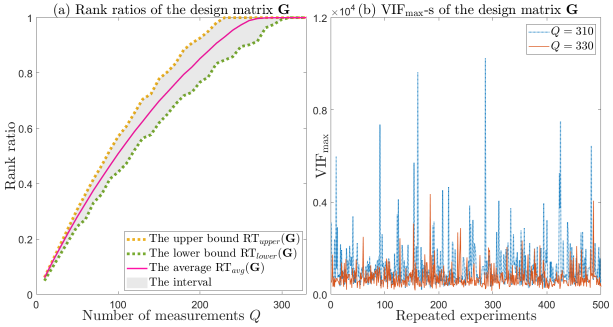

Figure 2: Statistical characteristics of the design matrix under AES ($n = 8$) and $d = 4$: (a): rank ratios; (b): VIF$_{\max}$-s.

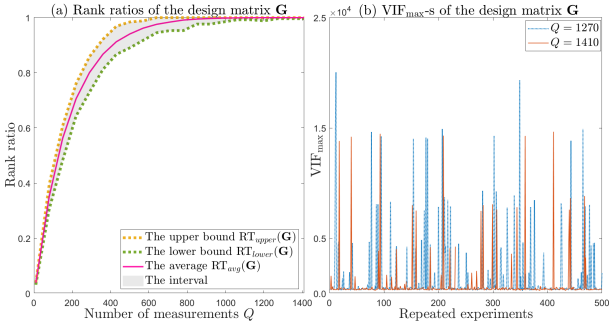

Figure 3: Statistical characteristics of the design matrix under AES ($n = 8$) and $d = 7$: (a): rank ratios; (b): VIF$_{\max}$-s.

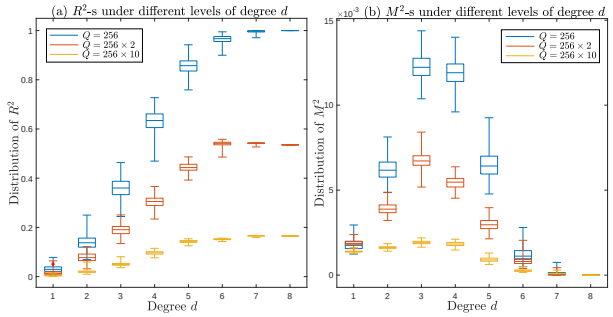

Figure 4: Distributions of R^2 and M^2 for AES ($n = 8$) under different guessing subkeys k and increasing degrees d.

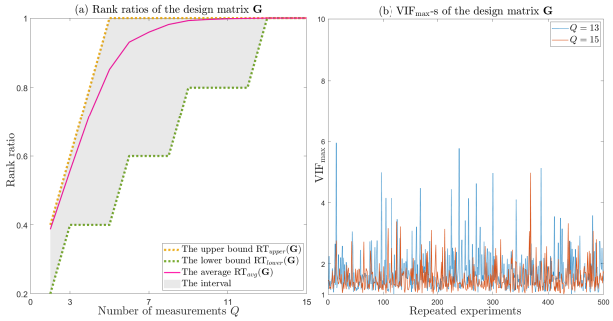

Figure 5: Statistical characteristics of the design matrix under PRESENT ($n = 4$) and $d = 1$: (a): rank ratios; (b): VIF$_{\max}$-s.

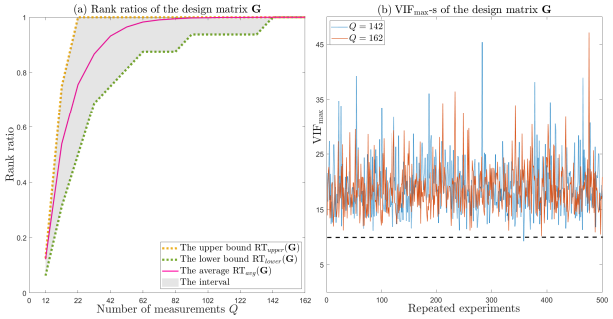

Figure 6: Statistical characteristics of the design matrix under PRESENT ($n = 4$) and $d = 4$: (a): rank ratios; (b): VIF$_{\max}$-s.

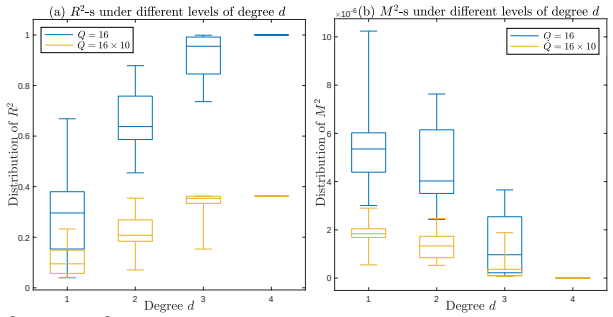

Figure 7: Distributions of R^2 and M^2 for PRESENT ($n = 4$) under different guessing subkeys k and increasing degrees d.

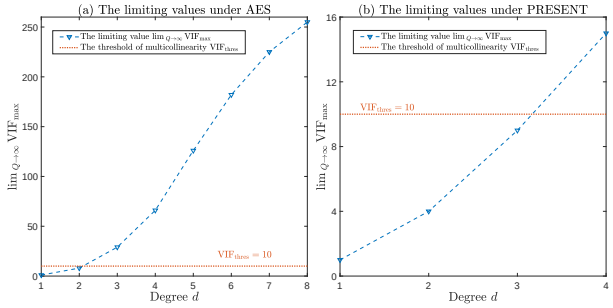

Figure 8: The limiting values of VIF$_{\max}$ in AES and PRESENT.

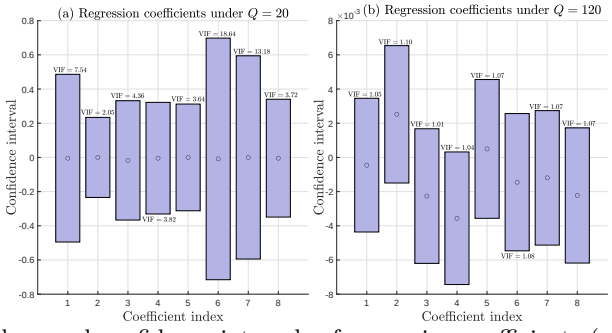

Figure 9: The estimated values and confidence intervals of regression coefficients ($\alpha = 0.95$): (a): $Q = 20$; (b): $Q = 120$.

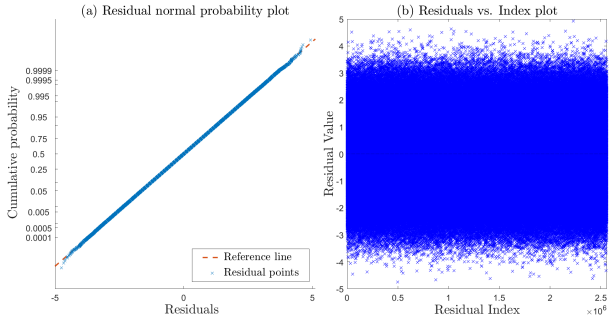

Figure 10: The statistical characteristics of residuals on HW leakages with $k = k^*$.

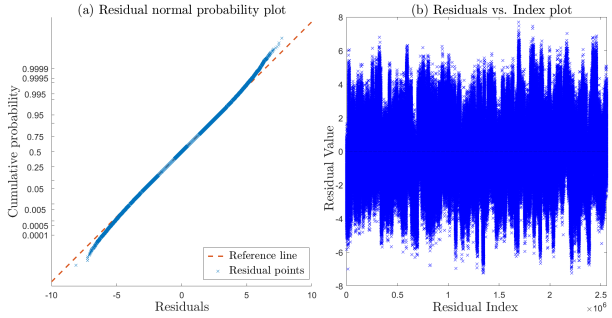

Figure 11: The statistical characteristics of residuals on HW leakages with $k \neq k^*$.

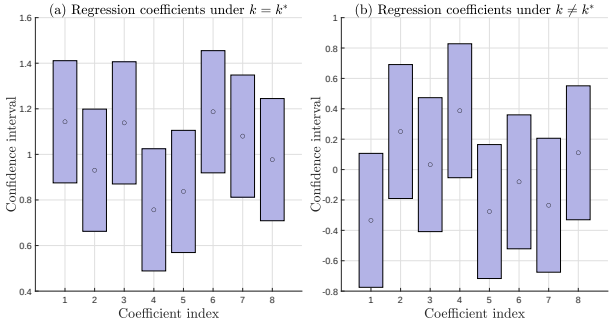

Figure 12: The estimated values and confidence intervals of regression coefficients ($\alpha = 0.95$): (a): $k = k^*$; (b): $k \neq k^*$.

5.2 Experimental Analyses on Residuals

Experimental results of the second experiment are given in Figs. 10∼12. We investigate the most common leakage $\mathcal{L} = \mathrm{HW}(\mathrm{Sbox}(m \oplus k^*)) + \mathsf{N}$ of AES and ensure both conditions in **Theorem 2** are met. This allows for a systematical analysis solely from the aspect of residuals, excluding the possible effects of multicollinearity and potential explanatory variables that are not considered. Our observations are as follows:

1. Residuals deviate from normality when the subkey is incorrect. In Figs. 10(a)∼11(a), we intuitively compare the statistical distributions of residuals under different settings of subkey. The x-axis represents the residual values and the y-axis shows their corresponding cumulative probabilities. A red reference line indicating the theoretical normal distribution is given. As depicted, the cumulative distribution of residuals under $k = k^*$ always aligns well with the normal distribution; however, significant deviations occur at both ends of the reference line when the subkey is incorrect $k \neq k^*$, indicating issues of improper skewness and kurtosis in the distribution.

2. Residuals deviate from homoscedasticity and independence when the subkey is incorrect. In Figs. 10(b)∼11(b), we display the statistical distributions of residuals as the encryption progresses. The x-axis represents the indices of residuals (i.e., the order of appearance during encryptions) and the y-axis shows the residual values. Under $k = k^*$, residuals are uniformly distributed along with the encryption process, adhering to homoscedasticity and independence. However, they become erratic and volatile given an incorrect subkey. Together with the above findings, these two observations corroborate **Lemma 3**.

3. LRA is advantageous for evaluators. In Fig 12 we compare the estimated coefficients of evaluators and adversaries who have equivalent strength, i.e., based on entirely identical measurements and degree. The only advantage of evaluators is the knowledge of subkey. As illustrated, excluding the effects of multicollinearity, confidence intervals are equal in length for all explanatory variables under the same subkey. It is noteworthy that the estimated results are more accurate for evaluators under $k = k^*$, as evidenced by their confidence intervals being shorter than those of $k \neq k^*$. Naturally, this will facilitate better statistical tests or model building in further applications of LRA. This observation corroborates **Theorem 3**.

5.3 Experimental Analyses on Estimated Outcomes

Experimental results of the third experiment are provided in Figs. 13∼15. To validate the asymptotic convergence behaviors of the regression outcomes, we choose the Sum of Absolute Errors (SAE) as the evaluation metric. The summation is performed over subkeys to fully verify the theorems, i.e., errors for all subkeys should reduce to zero. By Equ.(25), for the regression coefficients, $\mathrm{SAE}(\beta_i) = \sum_{k\in\mathcal{K}} |\beta_i(k) - \Delta^{(k)}_{\mathsf{SB-DoM}}(i)|$. For the regression intercept, $\mathrm{SAE}(\beta_{-1}) = \sum_{k\in\mathcal{K}} |\beta_{-1}(k) - (\mathbb{E}\{\mathcal{L}_q\} - \frac{1}{2}\Delta^{(k)}_{\mathsf{M-DoM}})|$. This verification will also directly support Equ.(27). By Equ.(29), for the coefficient of determination, $\mathrm{SAE}(R^2) = \sum_{k\in\mathcal{K}} |R^2(k) - \rho_{adj}(k)|$. The term ρ_{adj} is the adjusted version by PVE. For comparison we also provide results of the original ρ. Our observations are:

1. **Theorem 4** and **5** for LRA hold true in practice. In Fig. 13(a) for AES, under any given Q, there are a total of eight regression coefficients (plotted as blue crosses) plus one intercept (plotted as a red cross). As the number of measurements increase, the SAE approaches zero rapidly. Similar phenomena are observed in Fig. 14(a) for PRESENT (four regression coefficients plus one intercept) where the higher SNR leads to perfect verification with much smaller SAE-s. It is noteworthy that the red crosses representing the intercepts have higher SAE values in overall. This is mainly because they relate to the differential values of M-DoM which are sums of those from SB-DoM, thereby aggregating their errors.

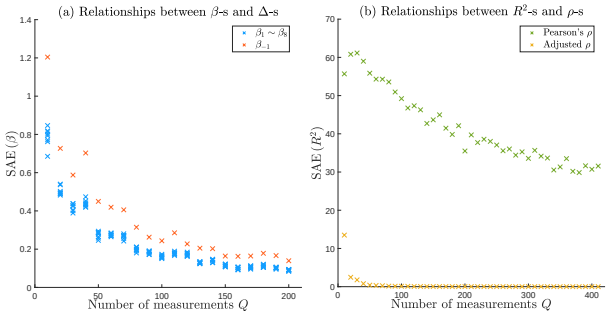

Figure 13: Relationships between LRA and traditional techniques (AES): (a): β and Δ; (b): R^2 and ρ.

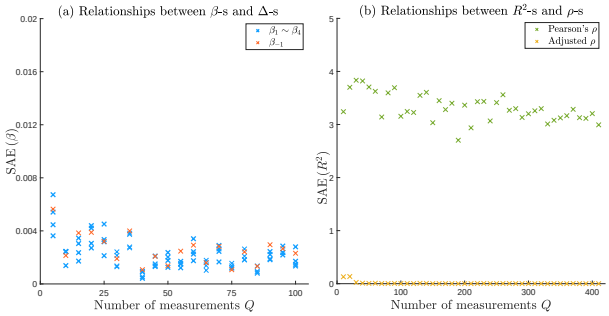

Figure 14: Relationships between LRA and traditional techniques (PRESENT): (a): β and Δ; (b): R^2 and ρ.

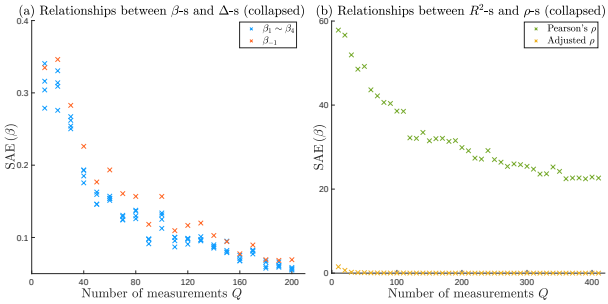

Figure 15: Collapsed function enhanced LRA and its relationships with G-DoM (AES): (a): β and Δ; (b): R^2 and ρ.

2. **Theorem 6** for LRA holds true in practice. In Fig. 13(b)∼14(b), the green crosses represent the original Pearson's correlation coefficients and the brown ones indicate the adjusted versions by PVE. For the adjusted ρ, for both AES and PRESENT, the instantaneous decays of SAE to zero complete our verification. However, the SAE-s under the Pearson's ρ maintain high levels and tend to constant under large measurement set. From distinguisher's perspectives, LRA and CPA achieve similar performance with high success rates in these cases. So, there are only constant differences between their distinguishing values which do not affect the attack result.

3. **Theorem 7** for LRA holds true in practice. In the experiment, we arbitrarily divide the binary bits of the intermediate variable into four bit groups of varying length: $\{\{x^{[5]}\}, \{x^{[2]}\}, \{x^{[1]}, x^{[7]}\}, \{x^{[6]}, x^{[8]}, x^{[3]}\}\}$. We discard the fourth bit $x^{[4]}$. This allows a robust verification of the theorem. Let $\{g_1, g_2, g_3, g_4\}$ denote the bit groups. Under this setting, the SAE-s for the intercept and the four regression coefficients corresponding to the collapsed outputs are calculated as: $\mathrm{SAE}(\beta_i) = \sum_{k\in\mathcal{K}} |\beta_i(k) - \Delta^{(k)}_{\mathsf{G-DoM}}(g_i)|$. For the regression intercept, $\mathrm{SAE}(\beta_{-1}) = \sum_{k\in\mathcal{K}} |\beta_{-1}(k) - (\mathbb{E}\{\mathcal{L}_q\} - \frac{1}{2}\sum_i \Delta^{(k)}_{\mathsf{G-DoM}}(g_i))|$. As shown in Fig. 15, again, the small SAE-s verify the correctness and robustness of this theorem.

6 Conclusions

This paper centers around LRA within the field of side-channel. We address the fundamental problems of why and how to use LRA and theoretically elaborate its connections with traditional side-channel techniques. We also study the state-of-the-art combined LRA with the collapse functions. In the future, we will extend our research to more advanced scenarios such as ridge and Lasso leakage regression [15].

References

[1] Si Gao and Elisabeth Oswald. A novel completeness test for leakage models and its application to side channel attacks and responsibly engineered simulators. In EUROCRYPT 2022, Trondheim, Norway, May 30 - June 3, 2022, Proceedings, Part III, pages 254–283. Springer, 2022.

[2] Si Gao, Elisabeth Oswald, and Dan Page. Towards micro-architectural leakage simulators: Reverse engineering micro-architectural leakage features is practical. In EUROCRYPT 2022, Trondheim, Norway, May 30 - June 3, 2022, Proceedings, Part III, pages 284–311. Springer, 2022.

[3] Si Gao and Elisabeth Oswald. A novel framework for explainable leakage assessment. In EUROCRYPT 2024, Zurich, Switzerland, May 26-30, 2024, Proceedings, Part III, pages 221–250. Springer, 2024.

[4] Stefan Mangard, Elisabeth Oswald, and Thomas Popp. Power analysis attacks: Revealing the secrets of smart cards, volume 31. Springer Science & Business Media, 2008.

[5] Dakshi Agrawal, Bruce Archambeault, Josyula R Rao, and Pankaj Rohatgi. The em side—channel (s). In CHES 2002, CA, USA, August 13–15, 2002 Revised Papers 4, pages 29–45. Springer, 2003.

[6] Common Criteria: The Common Criteria for Information Technology Security Evaluation (2017). https://www.commoncriteriaportal.org/cc/.

[7] Information Technology Laboratory, NIST: Security Requirements for Cryptographic Modules. https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.140-3.pdf.

[8] François-Xavier Standaert, Tal G Malkin, and Moti Yung. A unified framework for the analysis of side-channel key recovery attacks. In EUROCRYPT 2009, Cologne, Germany, April 26-30, 2009. Proceedings 28, pages 443–461. Springer, 2009.

[9] Eloi De Chérisey, Sylvain Guilley, Olivier Rioul, and Pablo Piantanida. Best information is most successful. Cryptology ePrint Archive, 2019.

[10] Akira Ito, Rei Ueno, and Naofumi Homma. On the success rate of side-channel attacks on masked implementations: information-theoretical bounds and their practical usage. In Proceedings of the 2022 ACM SIGSAC Conference on CCS, pages 1521–1535, 2022.

[11] Benjamin Jun Gilbert Goodwill, Josh Jaffe, Pankaj Rohatgi, et al. A testing methodology for side-channel resistance validation. In NIST non-invasive attack testing workshop, volume 7, pages 115–136, 2011.

[12] François Durvaux and François-Xavier Standaert. From improved leakage detection to the detection of points of interests in leakage traces. In EUROCRYPT 2016, Vienna, Austria, May 8-12, 2016, Proceedings, Part I 35, pages 240–262. Springer, 2016.

[13] Amir Moradi, Bastian Richter, Tobias Schneider, and François-Xavier Standaert. Leakage detection with the x2-test. IACR Transactions on CHES, pages 209–237, 2018.

[14] Carolyn Whitnall, Elisabeth Oswald, and François-Xavier Standaert. The myth of generic dpa… and the magic of learning. In CT-RSA 2014, San Francisco, CA, USA, February 25-28, 2014. Proceedings, pages 183–205. Springer, 2014.

[15] Weijia Wang, Yu Yu, Junrong Liu, Zheng Guo, François-Xavier Standaert, Dawu Gu, Sen Xu, and Rong Fu. Evaluation and improvement of generic-emulating dpa attacks. In CHES 2015, Saint-Malo, France, September 13-16, 2015, Proceedings 17, pages 416–432. Springer, 2015.

[16] Suresh Chari, Josyula R Rao, and Pankaj Rohatgi. Template attacks. In CHES 2002, CA, USA, August 13–15, 2002 Revised Papers 4, pages 13–28. Springer, 2003.

[17] Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual information analysis: how, when and why? In International Workshop on CHES, pages 429–443. Springer, 2009.

[18] Benedikt Gierlichs, Lejla Batina, Pim Tuyls, and Bart Preneel. Mutual information analysis: A generic side-channel distinguisher. In International Workshop on CHES, pages 426–442. Springer, 2008.

[19] François Durvaux, François-Xavier Standaert, and Nicolas Veyrat-Charvillon. How to certify the leakage of a chip? In EUROCRYPT 2014, Copenhagen, Denmark, May 11-15, 2014. Proceedings 33, pages 459–476. Springer, 2014.

[20] P Kocher. Differential power analysis. In Proc. Advances in Cryptology (CRYPTO'99), 1999.

[21] Eric Brier, Christophe Clavier, and Francis Olivier. Correlation power analysis with a leakage model. In CHES 2004, MA, USA, August 11-13, 2004. Proceedings 6, pages 16–29. Springer, 2004.

[22] Werner Schindler, Kerstin Lemke, and Christof Paar. A stochastic model for differential side channel cryptanalysis. In CHES 2005, Edinburgh, UK, August 29–September 1, 2005. Proceedings 7, pages 30–46. Springer, 2005.

[23] Ralph B D'Agostino. Goodness-of-fit-techniques. Routledge, 2017.

[24] Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning, volume 4. Springer, 2006.

[25] David McCann, Elisabeth Oswald, and Carolyn Whitnall. Towards practical tools for side channel aware software engineering:'grey box'modelling for instruction leakages. In 26th USENIX security symposium (USENIX Security 17), pages 199–216, 2017.

[26] Julien Doget, Emmanuel Prouff, Matthieu Rivain, and François-Xavier Standaert. Univariate side channel attacks and leakage modeling. Journal of Cryptographic Engineering, 1:123–144, 2011.

[27] Sylvain Guilley, Annelie Heuser, and Olivier Rioul. A key to success: Success exponents for side-channel distinguishers. In International Conference on Cryptology in India, pages 270–290. Springer, 2015.

[28] Stefan Mangard, Elisabeth Oswald, and F-X Standaert. One for all–all for one: unifying standard differential power analysis attacks. IET Information Security, 5(2):100–110, 2011.

[29] Régis Bevan and Erik Knudsen. Ways to enhance differential power analysis. In International Conference on Information Security and Cryptology, pages 327–342. Springer, 2002.

[30] Amir Moradi and François-Xavier Standaert. Moments-correlating dpa. In Proceedings of the 2016 ACM Workshop on Theory of Implementation Security, pages 5–15, 2016.

[31] Thanh-Ha Le, Jessy Clédière, Cécile Canovas, Bruno Robisson, Christine Servière, and Jean-Louis Lacoume. A proposition for correlation power analysis enhancement. In CHES 2006, Yokohama, Japan, October 10-13, 2006. Proceedings 8, pages 174–186. Springer, 2006.

[32] Mathieu Renauld, François-Xavier Standaert, Nicolas Veyrat-Charvillon, Dina Kamel, and Denis Flandre. A formal study of power variability issues and side-channel attacks for nanoscale devices. In EUROCRYPT 2011, Tallinn, Estonia, May 15-19, 2011. Proceedings 30, pages 109–128. Springer, 2011.

[33] Wilhelm Forst and Dieter Hoffmann. Optimization—theory and practice. Springer Science & Business Media, 2010.

[34] David A Belsley, Edwin Kuh, and Roy E Welsch. Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons, 2005.

[35] James Durbin and Geoffrey S Watson. Testing for serial correlation in least squares regression. iii. Biometrika, 58(1):1–19, 1971.

[36] Stanley Smith Stevens. On the theory of scales of measurement. Science, 103(2684):677–680, 1946.

[37] John Neter, Michael H Kutner, Christopher J Nachtsheim, William Wasserman, et al. Applied linear statistical models. 1996.

[38] 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, 2011.

[39] Jiangshan Long, Changhai Ou, Chenxu Wang, Zhu Wang, and Yongbin Zhou. What is now possible? security evaluation on univariate dpa attacks with inaccurate leakage models. IEEE Transactions on Information Forensics and Security, 2024.

[40] DPA contest v4.1. https://www.dpacontest.org/home/. Accessed 2024.5.13.