A Statistical Model for DPA with Novel Algorithmic Confusion Analysis

Yunsi Fei[1], Qiasi Luo[2,⋆], and A. Adam Ding[3]

[1] Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115
[2] Marvell Technology Group Ltd., Santa Clara, CA 95054
[3] Department of Mathematics, Northeastern University, Boston, MA 02115

Abstract. Side-channel attacks (SCAs) exploit weakness in the physical implementation of cryptographic algorithms, and have emerged as a realistic threat to many critical embedded systems. However, no theoretical model for the widely used differential power analysis (DPA) has revealed exactly what the success rate of DPA depends on and how. This paper proposes a statistical model for DPA that takes characteristics of both the physical implementation and cryptographic algorithm into consideration. Our model establishes a quantitative relation between the success rate of DPA and a cryptographic system. The side-channel characteristic of the physical implementation is modeled as the ratio between the difference-of-means power and the standard deviation of power distribution. The side-channel property of the cryptographic algorithm is extracted by a novel algorithmic confusion analysis. Experimental results on DES and AES verify this model and demonstrate the effectiveness of algorithmic confusion analysis. We expect the model to be extendable to other SCAs, and provide valuable guidelines for truly SCA-resilient system design and implementation.

Keywords: Side-channel attack, differential power analysis.

1 Introduction

Cryptographic algorithms are widely used in various computer systems to ensure security. Despite the security strength of the algorithm, the leaked side-channel information of the cryptosystem implementation, like power consumption of smart cards and timing information of embedded processors, can be exploited to recover the secret key. Differential power analysis (DPA) is one of the early effective SCAs which analyzes the correlation between intermediate data and power consumption to reveal the secret [1]. Over the past decade, there has been many other successful power analysis attacks, including Correlation Power Attack (CPA) [2], Mutual Information Attack (MIA) [3], Partitioning Power Analysis (PPA) [4], etc. Other side-channel information, like electromagnetic emanations [5,6] and timing information [7], can also be exploited. A real secure

⋆ This work was done while the author was with University of Connecticut.

E. Prouff and P. Schaumont (Eds.): CHES 2012, LNCS 7428, pp. 233–250, 2012.

system must be designed with countermeasures to be SCA-resilient. Common countermeasures include masking [8], power-balanced logic [9], and random delays [10]. To measure the SCA resilience of a system or the effectiveness of a countermeasure, several generic metrics are used, such as *number of measurements*, *success rate* [11,12], *guessing entropy* [13] and *information theoretic metric* [13,14]. One commonly used metric for evaluating a system's SCA resilience is the success rate, i.e., the probability that a specific SCA is successful with certain complexity constraint. For a cryptosystem, a low success rate for a SCA on it indicates its high resilience against such SCA.

Intuitively, both the *physical implementation* and *cryptographic algorithm* would affect the SCA resilience of a cryptosystem. An ideal implementation with countermeasures could reduce the side-channel leakage to minimum. Different cryptographic algorithms may have different intrinsic SCA-related properties. Accurately evaluating different implementations of the same cryptographic algorithm and comparing different cryptographic algorithms, in terms of their SCA resilience, are challenging issues. However, such analysis and theoretical modeling will reveal system-inherent parameters that affect its SCA resilience, and in practice will greatly facilitate advances in the design and implementation of real secure cryptosystems.

Related Work. There has been many related research efforts attempting to address the above issues. However, the effects of the algorithm and implementation were not clearly decoupled and better quantitative model is needed to understand their interaction. In [15], an approach is presented to model the DPA signal-to-noise ratio (SNR) of a cryptographic system, which does not further reveal how the SNR determines the ultimate SCA resilience. In [16], the relation between the difference-of-means power consumption and key hypotheses is analyzed and utilized to improve the DPA efficiency, without examining characteristics of the algorithm. [17] presented a statistical model for CPA, which illustrated well the effect of SNR on the power of CPA. However, they did not consider the interaction between the incorrect keys and thus the formula does not numerically conform to the empirical overall success rate for CPA (see Appendix A). Work in [18] exhibits DPA-related properties of SBoxes in cryptographic algorithms and introduces a new notion of *transparency order of an SBox*, without considering the implementation aspect. A framework presented in [13] unifies the theory and practice of SCA with a combination of information theory and security metrics. A quantitative analysis between the metrics and cryptographic system would be a nice complement to the general framework.

Our Contributions. In this paper, we proposes a statistical analysis model for DPA. To the best of our knowledge, this is the first analytic model for the success rate of DPA on cryptographic systems, and also the first model extracting SCA related characteristics from both the physical implementation and cryptographic algorithm. The physical implementation is represented by the power difference related to the select function and standard deviation of power waveforms. The ratio between them defines the SCA resilience of an implementation. The SCA-related property of a cryptographic algorithm is characterized by algorithmic

confusion analysis. A confusion matrix is generated to measure the statistical correlation between different key candidates in DPA.

The rest of the paper is organized as follows. Section 2 introduces notions and fundamentals in cryptographic algorithms, DPA procedures, and statistical aspects in SCAs. Section 3 presents the algorithmic confusion analysis with definitions of confusion and collision coefficients. Our model for the success rate of DPA is proposed in Section 4. The model is verified with experimental results on DES and AES in Section 5. Section 6 discusses more implications of the model and its possible applications. Finally conclusions are drawn in Section 7.

2 Preliminaries

2.1 Randomness of Cryptographic Algorithm

Cryptographic algorithms are designed to be robust against cryptanalysis with two well-known statistical properties [19]. *Confusion* makes the statistical relation between the the ciphertext and key as complex as possible; *diffusion* makes the statistical relation between the ciphertext and plaintext as complex as possible. With deliberate design, an encryption algorithm is *perfectly secret* if each bit in the ciphertext C is purely random [20]:

Theorem 1. *Suppose b_C is one bit of the ciphertext C for a perfectly secret encryption algorithm, b_C has the same probability to be* 0 *or* 1*:*

$$\Pr[b_C = 1] = \Pr[b_C = 0] = \frac{1}{2}.$$

2.2 Differential Power Analysis Procedure

All SCAs have a common hypothesis test procedure. We next give an introduction on the earliest discovered and important DPA procedure.

- Side-channel *measurements* obtain physical side-channel information W, i.e., waveforms of power consumption collected from devices. Denote the *waveform population* as $\mathcal{W} = \{W_1, \ldots, W_{N_m}\}$, where W_i is a (time series) measurement with a certain input, and N_m is the total number of measurements for the cryptographic system. Each W is a time series as $W = \{w^1, \ldots, w^{N_p}\}$, where N_p is the number of points in W.
- *Key hypotheses* enumerate all possible values of the subkey k under attack, denoted as $\langle k \rangle = \{k_0, \ldots, k_{N_k-1}\}$, where N_k is the total number of key guesses, and $N_k = 2^{l_k}$ with l_k as the subkey bit-length.
- *Select function* ψ for DPA is one single bit b_d of *intermediate data* d computed from the plaintext M or ciphertext C and a key, written as $\psi = b_d$. The value of ψ is either 1 or 0.
- *Correlation* between ψ for each key hypothesis and $\mathcal{W}$ is computed for a specific attack. The correlation for DPA is the difference of means (DoM) δ,

i.e., the difference between the average power consumption of two waveform groups partitioned with $\psi = 1$ and 0, written as:

$$\delta = \frac{\sum \mathcal{W}_{\psi=1}}{N_{\psi=1}} - \frac{\sum \mathcal{W}_{\psi=0}}{N_{\psi=0}} \tag{1}$$

where $N_{\psi=1}$ and $N_{\psi=0}$ are the numbers of measurements with $\psi = 1$ and $\psi = 0$ respectively, under a particular key hypothesis. Given enough number of measurements, the DoM δ_c for the correct key k_c converges to the power difference ϵ related to the bit b_d under attack, written as $\lim_{N_m \to \infty} \delta_c = \epsilon$, where $N_m = N_{\psi=1} + N_{\psi=0}$.

– *Testing* with the maximum likelihood method chooses the key hypothesis which has the maximum correlation (DoM in DPA) as the correct key.

2.3 Central Limit Theorem

The basic statistical aspect of our model is the Central Limit Theorem [21], considering the various noises in leakage measurements and the sampling process for side-channel cryptanalysis, i.e., the leakage measurement is for a set of random inputs rather than enumerating the entire input space. Consider a random distribution $\mathcal{X} = \{x_1, x_2, x_3, \ldots\}$. The mean value and standard deviation of the population are μ and σ, respectively. Randomly select a sample of size N_x from the population we get the mean value:

$$\bar{X} = \frac{1}{N_x} \sum_{i=1}^{N_x} x_i.$$

When N_x is sufficiently large, $\bar{X}$ is approximately normally distributed, $\mathcal{N}(\mu_{\bar{X}}, \sigma_{\bar{X}})$, with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N_x}}$.

DPA is a sampling process on the entire waveform population, which is usually regarded as normally distributed [22]. Denote the standard deviation of the waveform population as $\sigma_{\mathcal{W}}$. Thus the two mean terms for the DoM computation in Equation (1) are normal random variables with distribution $\mathcal{N}\left(\epsilon + b, \sigma_{\mathcal{W}}/\sqrt{N_{\psi=1}}\right)$ and $\mathcal{N}\left(b, \sigma_{\mathcal{W}}/\sqrt{N_{\psi=0}}\right)$, respectively. Here b denotes the mean power consumption for the waveform group $\psi = 0$. Since both $N_{\psi=0}$ and $N_{\psi=1}$ are approximately $\frac{N_m}{2}$ according to Theorem 1, δ_c is a random variable with normal distribution $\mathcal{N}\left(\mu_{\delta_c}, \sigma_{\delta_c}\right)$ as $\mu_{\delta_c} = \epsilon$ and $\sigma_{\delta_c} = 2\frac{\sigma_{\mathcal{W}}}{\sqrt{N_m}}$. This statement still holds for large N_m by the Central Limit Theorem when we drop the normal distribution assumption on the waveform population.

3 Algorithmic Confusion Analysis

A chosen select function involves a certain SBox of the cryptographic algorithm (a preset computation given as a lookup table) and a subkey. In this section, we attempt to reveal properties of the algorithm that would indicate its resilience to DPA. The analysis is only algorithm and select function related, and independent on the leakage measurements.

3.1 Confusion Coefficient

Assume the select function for DPA is chosen as a bit in the last-round encryption, which is dependent on several bits of the ciphertext, the subkey, and the corresponding SBox. Two key hypotheses k_i and k_j would have two corresponding $\psi|k_i$ and $\psi|k_j$. The values of $\psi|k_i$ and $\psi|k_j$ can be different or the same. We find out that the probability that $\psi|k_i$ is different or the same with $\psi|k_j$ reveals DPA-related property of the cryptographic algorithm.

We name a *confusion coefficient* after the confusion property of cryptographic algorithms defined in [19]. The *confusion coefficient* κ over two keys (k_i, k_j) is defined as:

$$\kappa = \kappa(k_i, k_j) = \Pr\left[(\psi|k_i) \neq (\psi|k_j)\right] = \frac{N_{(\psi|k_i)\neq(\psi|k_j)}}{N_t}$$

where N_t is the total number of values for the relevant ciphertext bits, and $N_{(\psi|k_i)\neq(\psi|k_j)}$ is the number of occurrences for which different key hypotheses k_i and k_j result in different ψ values. For example, in our DPA attack on DES (Data Encryption Standard) algorithm, N_t is $2^7 = 128$.

Similarly, the *complementary confusion coefficient* or *collision coefficient* ξ over (k_i, k_j) is defined as:

$$\xi = \xi(k_i, k_j) = \Pr\left[(\psi|k_i) = (\psi|k_j)\right] = \frac{N_{(\psi|k_i)=(\psi|k_j)}}{N_t}$$

We have $\kappa + \xi = 1$ and $0 \leq \kappa < 1$ and $0 < \xi \leq 1$. For a perfectly secret cryptographic, we have:

Lemma 1. *Confusion Lemma (see Appendix B for the proof).*

$$\Pr\left[(\psi|k_i) = 0, (\psi|k_j) = 1\right] = \Pr\left[(\psi|k_i) = 1, (\psi|k_j) = 0\right] = \frac{1}{2}\kappa$$
$$\Pr\left[(\psi|k_i) = 1, (\psi|k_j) = 1\right] = \Pr\left[(\psi|k_i) = 0, (\psi|k_j) = 0\right] = \frac{1}{2}\xi.$$

For three different keys k_h, k_i and k_j, we further define a three-way confusion coefficient:

$$\tilde{\kappa} = \tilde{\kappa}(k_h, k_i, k_j) = \Pr\left[(\psi|k_i) = (\psi|k_j), \psi|k_i) \neq (\psi|k_h)\right].$$

Lemma 2. $\tilde{\kappa}(k_h, k_i, k_j) = \frac{1}{2}[\kappa(k_h, k_i) + \kappa(k_h, k_j) - \kappa(k_i, k_j)]$. *(See Appendix C)*

3.2 Confusion Coefficient and DPA

The power measurements are for one key embedded in the cryptographic system, i.e., the correct key, denote as k_c. Denote k_g as one of the incorrect key guesses. Suppose the DoM for k_c and k_g are δ_c and δ_g, respectively. The difference between the two DoMs is $\Delta(k_c, k_g) = (\delta_c - \delta_g)$. We have obtained the mean and variance of $\Delta(k_c, k_g)$ (see Appendix D):

$$\begin{aligned} E[\Delta(k_c, k_g)] &= 2\kappa(k_c, k_g)\epsilon \\ Var[\Delta(k_c, k_g)] &= 16\kappa(k_c, k_g)\frac{\sigma_{\mathcal{W}}^2}{N_m} + 4\kappa(k_c, k_g)\xi(k_c, k_g)\frac{\epsilon^2}{N_m} \end{aligned} \tag{2}$$

Hence, $\lim_{N_m \to \infty} \Delta(k_c, k_g) = 2\kappa(k_c, k_g)\epsilon$.

4 Statistical Model for DPA

In DPA, to successfully distinguish the correct key k_c from other key hypotheses, it requires the DoM of k_c to be larger than that of all other keys, written as: $\delta_{k_c} > \{\delta_{\langle \overline{k_c} \rangle}\}$, where $\langle \overline{k_c} \rangle$ denotes all the incorrect keys, i.e., $\{k_0, \ldots, k_{N_k-1}\}$ excluding k_c, and $\{\delta_{\langle \overline{k_c} \rangle}\}$ denotes $\{\delta_{k_0}, \ldots, \delta_{k_{N_k-1}}\}$ excluding δ_{k_c}. The success rate to recover the correct key, SR, is defined as the probability for $\delta_{k_c} > \delta_{\langle \overline{k_c} \rangle}$:

$$\mathrm{SR} = \mathrm{SR}\left[k_c, \langle \overline{k_c} \rangle\right] = \Pr\left[\delta_{k_c} > \{\delta_{\langle \overline{k_c} \rangle}\}\right]$$

The overall success rate is against $(N_k - 1)$ wrong keys. We next show the derivation of the success rates starting from the simple one-key success rate.

1-key Success Rate. We first consider the 1-key success rate, i.e., the success rate of k_c over an incorrect key k_g chosen out of $\langle \overline{k_c} \rangle$, written as:

$$\mathrm{SR}_1 = \mathrm{SR}\left[k_c, k_g\right] = \Pr\left[\delta_{k_c} > \delta_{k_g}\right] = \Pr\left[\Delta(k_c, k_g) > 0\right].$$

From Section 2.3, $\Delta(k_c, k_g)$ is the difference of two normal random variables, therefore follows distribution $\mathcal{N}\left(\mu_{\Delta(k_c,k_g)}, \sigma_{\Delta(k_c,k_g)}\right)$. From Equation (2),

$$\mu_{\Delta(k_c,k_g)} = 2\kappa(k_c, k_g)\epsilon, \qquad \sigma_{\Delta(k_c,k_g)} = 2\sqrt{\frac{\kappa(k_c, k_g)}{N_m}}\sqrt{4\sigma_{\mathcal{W}}^2 + \xi(k_c, k_g)\epsilon^2}.$$

Let $\Phi(x) = \frac{1}{2}[1 + erf(\frac{x}{\sqrt{2}})]$ denote the cumulative distribution function (cdf) of the standard normal distribution, where erf(x) is the error function $erf(x) = \frac{2}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$. Since $\frac{\Delta(k_c,k_g)-\mu_{\Delta(k_c,k_g)}}{\sigma_{\Delta(k_c,k_g)}}$ is a standard normal random variable,

$$\begin{aligned}\mathrm{SR}_1 &= \Pr\left[\Delta(k_c, k_g) > 0\right] = 1 - \Phi(-\frac{\mu_{\Delta(k_c,k_g)}}{\sigma_{\Delta(k_c,k_g)}}) = \Phi(\frac{\mu_{\Delta(k_c,k_g)}}{\sigma_{\Delta(k_c,k_g)}}) \\ &= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\mu_{\Delta(k_c,k_g)}}{\sqrt{2}\sigma_{\Delta(k_c,k_g)}}\right)\right] = \frac{1}{2}\left[1 + \mathrm{erf}\left(\sqrt{\frac{\kappa(k_c, k_g)}{(\frac{2\sigma_{\mathcal{W}}}{\epsilon})^2 + \xi(k_c, k_g)}}\sqrt{\frac{N_m}{2}}\right)\right] \end{aligned} \tag{3}$$

This is a function of confusion coefficients $\kappa(k_c, k_g)$, the ratio of ϵ to $\sigma_{\mathcal{W}}$, and the number of measurements, N_m. Overall, the higher $\epsilon/\sigma_{\mathcal{W}}$, $\kappa(k_c, k_g)$, and N_m are, the higher the success rate is, i.e., more susceptible to DPA.

2-keys Success Rate. Next we consider the 2-keys success rate, i.e., the success rate of k_c over two chosen incorrect keys k_{g_1} and k_{g_2}, written as:

$$\mathrm{SR}_2 = \mathrm{SR}\left[k_c, \{k_{g_1}, k_{g_2}\}\right] = \Pr\left[\delta_{k_c} > \delta_{k_{g_1}}, \delta_{k_c} > \delta_{k_{g_2}}\right] = \Pr\left[y_1 > 0, y_2 > 0\right]$$

where

$$y_1 = \Delta(k_c, k_{g_1}) = \delta_{k_c} - \delta_{k_{g_1}}, \qquad y_2 = \Delta(k_c, k_{g_2}) = \delta_{k_c} - \delta_{k_{g_2}}.$$

Since y_1 and y_2 are random variables with normal distribution, $Y_2 = [y_1, y_2]^T$ is a random vector with two-dimension normal distribution as $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, where

$$\boldsymbol{\mu}_2 = \begin{bmatrix} \mu_{y_1} \\ \mu_{y_2} \end{bmatrix} = \begin{bmatrix} 2\kappa(k_c, k_{g_1})\epsilon \\ 2\kappa(k_c, k_{g_2})\epsilon \end{bmatrix}, \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} \mathrm{Cov}(y_1, y_1)\ \mathrm{Cov}(y_1, y_2) \\ \mathrm{Cov}(y_1, y_2)\ \mathrm{Cov}(y_2, y_2) \end{bmatrix}.$$

The covariances in $\boldsymbol{\Sigma}_2$ are (see Appendix E for the proof):

$$\begin{aligned}
\mathrm{Cov}(y_1, y_1) =& 16\kappa(k_c, k_{g_1})\frac{\sigma_{\mathcal{W}}^2}{N_m} + 4\kappa(k_c, k_{g_1})\xi(k_c, k_{g_1})\frac{\epsilon^2}{N_m} \\
\mathrm{Cov}(y_2, y_2) =& 16\kappa(k_c, k_{g_2})\frac{\sigma_{\mathcal{W}}^2}{N_m} + 4\kappa(k_c, k_{g_2})\xi(k_c, k_{g_2})\frac{\epsilon^2}{N_m} \\
\mathrm{Cov}(y_1, y_2) =& 16\tilde{\kappa}(k_c, k_{g_1}, k_{g_2})\frac{\sigma_{\mathcal{W}}^2}{N_m} + 4[\tilde{\kappa}(k_c, k_{g_1}, k_{g_2}) - \kappa(k_c, k_{g_1})\kappa(k_c, k_{g_2})]\frac{\epsilon^2}{N_m}.
\end{aligned}$$

Let $\Phi_2(\boldsymbol{x})$ denote the cdf of the 2-dimension standard normal distribution.

$$\mathrm{SR}_2 = \Phi_2(\sqrt{N_m}\boldsymbol{\Sigma}_2{}^{-1/2}\boldsymbol{\mu}_2) \tag{4}$$

which is a function of the ratio $\epsilon/\sigma_{\mathcal{W}}$, sample size N_m, and confusion coefficients $\kappa(k_c, k_{g_1})$, $\kappa(k_c, k_{g_2})$ and $\kappa(k_{g_1}, k_{g_2})$.

($\mathbf{N_k - 1}$)-keys success rate. The overall success rate is the success rate of k_c over all other $(N_k - 1)$ keys $\langle\overline{k_c}\rangle$,

$$\mathrm{SR} = \mathrm{SR}_{N_k-1} = \mathrm{SR}\left[k_c, \langle\overline{k_c}\rangle\right] = \Pr\left[\delta_{k_c} > \{\delta_{\langle\overline{k_c}\rangle}\}\right] = \Pr\left[Y > 0\right]$$

where Y is the $(N_k - 1)$-dimension vector of differences between δ_{k_c} and $\delta_{\langle\overline{k_c}\rangle}$:

$$Y = \delta_{k_c} - \delta_{\langle\overline{k_c}\rangle} = \left[\Delta(k_c, k_{g_1}), \ldots, \Delta(k_c, k_{g_{N_k-1}})\right]^T = [y_1, \ldots, y_{N_k-1}]^T.$$

Y is randomly distributed with $\mathcal{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$. The mean is:

$$\boldsymbol{\mu}_Y = 2\epsilon\boldsymbol{\kappa} \tag{5}$$

where $\boldsymbol{\kappa}$ denotes a $(N_k - 1)$-dimension *confusion vector* for the correct key k_c with entries $\kappa(k_c, k_{g_i})$, $i = 1, ..., N_k - 1$. The elements in the $(N_k - 1) \times (N_k - 1)$ matrix $\boldsymbol{\Sigma}_Y$ are covariances between $y_1, \ldots, y_{N_k-1}$. Thus

$$\boldsymbol{\Sigma}_Y = 16\frac{\sigma_{\mathcal{W}}^2}{N_m}\boldsymbol{K} + 4\frac{\epsilon^2}{N_m}(\boldsymbol{K} - \boldsymbol{\kappa}\boldsymbol{\kappa}^T) \tag{6}$$

where $\boldsymbol{\kappa}^T$ denotes the transpose of $\boldsymbol{\kappa}$, and $\boldsymbol{K}$ is the $(N_k - 1) \times (N_k - 1)$ *confusion matrix* of the cryptographic algorithm for k_c, with elements $\{\varkappa_{ij}\}$ as:

$$\varkappa_{ij} = \begin{cases} \kappa(k_c, k_{g_i}) & \text{if } i = j \\ \tilde{\kappa}(k_c, k_{g_i}, k_{g_j}) & \text{if } i \neq j. \end{cases}$$

The confusion matrix $\boldsymbol{K}$ fully depicts the relation between all the key candidates in the algorithm, and Equation (6) shows how it affects the success rate.

Let $\Phi_{N_k-1}(\boldsymbol{x})$ denote the cdf of the $(N_k - 1)$-dimension standard normal distribution.

$$\mathrm{SR} = \mathrm{SR}_{N_k-1} = \Phi_{N_k-1}(\sqrt{N_m}{\boldsymbol{\Sigma}_Y}^{-1/2}\boldsymbol{\mu}_Y). \tag{7}$$

Our statistical model for the overall success rate (SR) results in a multivariate Gaussian distribution formula. We can see that SR is determined by parameters related to both the physical implementation, ϵ and $\sigma_{\mathcal{W}}$, and the cryptographic algorithm, $\boldsymbol{K}$. ϵ and $\sigma_{\mathcal{W}}$ can be computed from the side-channel measurements of the cryptographic system. $\boldsymbol{K}$ is only determined by the specific selection function and cryptographic algorithm, independent of real physical implementations. Given these parameters, SR can be calculated with numerical simulations of the $(N_k - 1)$-dimension normal distribution. Our model extracts the effect of both the implementation and algorithm on SCA resilience quantitatively.

5 Experimental Results

5.1 DPA on DES

We perform DPA on DES, with the selection function on a randomly chosen bit. In our experiments, we choose the first bit of the input for the last round to evaluate the success rate model. We take the data set from DPAcontest [23] secmatv1 and focus on a single point (the 15750th point) which has the maximum DoM for k_c. Discussions on multi-point leakage will be given in Section 6.3. We generate the empirical success rate with 1000 trials as in [11,12].

To compute the theoretical success rate, we need the physical implementation parameters $SNR = \epsilon/\sigma_{\mathcal{W}}$ and the confusion coefficients κ for any two keys. Since k_c has been recovered for this data set, using *all* the power measurements at the selected leakage time point (the 15750th point), we can estimate ϵ as the DoM under k_c and estimate $\sigma^2_{\mathcal{W}}$ as the variance of power measurements. For a DES subkey of 6-bit, the number of key guesses, N_k, is 64, and there are (64×63) confusion coefficients. We found that they fall into nine values.

$$\{0.25, 0.3125, 0.375, 0.4375, 0.5, 0.5625, 0.625, 0.6875, 0.75\}.$$

We define these values as *characteristic confusion values* of a DES SBox. Why they end up in these nine values and what are the implications are unknown yet. However, we believe they manifest some important DPA-related properties of the SBoxes.

Fig. 1 plots the empirical success rates (the solid curves) and theoretical success rates (the dashed curves) of our model. We show the success rates against different number of key candidates for $k_c = k_{60}$. From top down, they are: $\mathrm{SR}_1 = \mathrm{SR}(k_c, k_0)$, $\mathrm{SR}_2 = \mathrm{SR}(k_c, \{k_0, k_1\})$, $\mathrm{SR}_8 = \mathrm{SR}(k_c, \{k_0, \ldots, k_7\})$, $\mathrm{SR}_{32} = \mathrm{SR}(k_c, \{k_0, \ldots, k_{31}\})$, and the overall success rate $\mathrm{SR}_{63} = \mathrm{SR}(k_c, \langle\overline{k_c}\rangle)$. We can see that the two curves for SR_{63} track each other very well, showing the accuracy of our theoretical model.

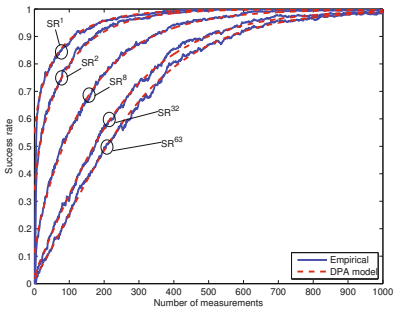

Fig. 1. Empirical and theoretical success rates of DPA on DES

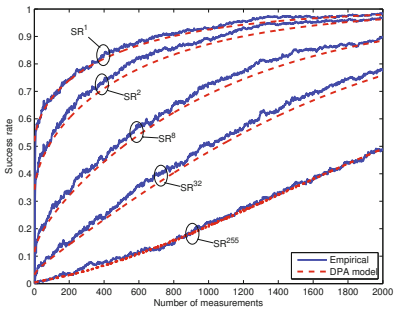

Fig. 2. Empirical and theoretical success rates of DPA on AES

5.2 DPA on AES

We next perform DPA on AES. The select function is defined as the XORed value of input and output of the third bit of the sixteenth SBox in the last round of AES. We measured the power consumption data using the SASEBO GII board with AES implementation designated by DPAcontest [24]. The total number of measurements in the data set is 100,000. The size of the AES subkey is 8, and there are (256×255) confusion coefficients, which also fall into nine characteristic confusion values of AES SBox:

$$\{0.4375, 0.453125, 0.46875, 0.484375, 0.5, 0.515625, 0.53125, 0.546875, 0.5625\}.$$

Fig. 2 shows the empirical success rates (solid curves) and theoretical success rates (dashed curves) of DPA for $k_c = k_{143}$. The two 255-keys success rate curves of empirical and theoretical track each other very well, demonstrating that the model is also very accurate for AES.

6 Discussions

Our DPA analysis builds a quantitative model for the SCA resilience of a cryptographic system over its inherent parameters, including ϵ, $\sigma_{\mathcal{W}}$ and $\boldsymbol{K}$. Next we present more SCA-related insights from the model about the implementation and algorithms, and how to use it to evaluate countermeasures and algorithms.

6.1 Signal and Noise of the Side Channel

Theoretically, DPA targets a portion of circuits that are related to the select function ψ, and other parts of the circuits are considered as random noise unrelated to ψ. DoM ϵ of the correct key is the power difference between $\psi = 1$ and $\psi = 0$ of the part of circuits under attack. DPA is a statistical process retrieving

the DoM ϵ out of all the power consumptions. As the number of power waveforms increases, the standard deviation of the difference between the DoMs for the correct key and incorrect keys decreases. When the standard deviations of DoMs are significantly less than ϵ, DPA has a significant success rate to recover the key. Here, ϵ indicates the signal level, and $\sigma_{\mathcal{W}}$ indicates the noise level.

We define $\epsilon/\sigma_{\mathcal{W}}$ as the *signal-to-noise ratio* (SNR) for the side channel. It is shown in Equation (3) for the 1-key success rate how the SNR determines the DPA results. The SNR can be used as a metric to measure the SCA resilience of the implementation of a cryptographic system. It is similar to the SNR defined in [15,22], however with more explicit quantitative implications in our model.

6.2 DPA-Confusion Property of Cryptographic Algorithms

Our algorithmic confusion analysis reveals the inherent property of a cryptographic algorithm, i.e., how differently the key candidates behave in DPA. Confusion coefficient is determined by both the cryptographic algorithm and selection function ψ. The eight different SBoxes in DES may have different confusion properties. Different bits in the same SBox may also have different confusion properties. Compared to DES, the confusion coefficients of an AES SBox are more concentrated near 0.5, which means the key candidates behave more randomly. For SBoxes with the same key space size, the success rates have the same dimensions, and hence the one with larger confusion coefficients leaks more information, leading to higher success rates. For two algorithms with different subkey space sizes, we need to compute the overall success rates. Comparing DES and AES, the dimension factor dominates over confusion coefficients. AES has 256 key candidates and the overall success rate is for 255-keys, making it more resilient than the 63-keys success rate of DES.

The experiments in Section 5.2 define the selection function ψ for DPA on AES as the XORed value of two intermediate data due to the characteristics of ASIC implementation. In micro-controller implementation, the select function is defined directly as one intermediate data. A good select function for attacks gives larger confusion coefficient $\kappa(k_c, k_g)$ and therefore larger success rate as shown in Equation (3). The algorithmic confusion analysis can also serve as a methodology to evaluate how good selection functions are at distinguishing the correct key.

6.3 Evaluation of DPA Countermeasure: Random Delay

Our model will be very useful for evaluating different DPA countermeasures. Here we take the method of random delay as example, which is an effective countermeasure to hide leakage [10,25,26]. We analyze the resilience of random delay under DPA to demonstrate the usage of our DPA model.

The random delay has no effect on the intermediate value. Thus it has no effect on the algorithmic confusion properties. It changes the success rate of DoM attack by affecting the signal-to-noise ratio. Random delay spreads out

the original side-channel leakages along the time, and therefore lowers the signal level. We consider the simplest case of single-point leakage at time s (general cases are presented in Appendix F). Suppose the power leakage without random delay is ϵ. The distribution of the random delay is $\Pr(t) = f_{rd}(t)$, for $t = 1, 2, ..., N_{rd}$ time units. Denote the time with largest $f_{rd}(t)$ as t_{max}. Then the maximum leakage after random shifting is $\varepsilon_{rd} = \varepsilon \cdot f_{rd}(t_{max})$, and the maximum leakage time point shifts from s to $s + t_{max}$. For uniform random delay, $\varepsilon_{rd} = \varepsilon / N_{rd}$. Larger N_{rd} would decrease ε_{rd}, however, it also slows down the program and degrades the performance. This also applies to more general multiple-point leakage. Our quantitative model can therefore aid the designer to fine-tune the balance between the SCA resilience and performance. Note that our success rate model is based on the knowledge of the correct key, k_c. It is meant to be adopted by the cryptosystem designer to take SCA security as a metric in the early design stage, by evaluating the SCA resilience of their implementations and countermeasures vigorously. It does not help the attacks.

6.4 Application of the Model to Other Side-Channel Attacks

Although there have been many other effective power analysis attacks, we choose the DPA to build the success rate model for its simplicity. As a matter of fact, all the power analysis attacks can be unified and it has been shown that the most popular approaches, such as DoM test, CPA, and Bayesian attacks, are essentially equivalent on a common target device (with the same power leakage model) [27]. DPA is the simplest one in modeling, because it targets a single bit. In CPA, the select function is the Hamming weight of the SBox output rather than a single bit. In addition, the correlation is the Pearson Correlation rather than the difference-of-means. We can envision that the success rate for CPA is still dependent on the implementation-determined parameters ϵ and $\sigma_{\mathcal{W}}$, and algorithm-dependent confusion coefficients κ and matrix. However, the confusion coefficient is no longer the probability that two different keys end up with different select function values, but would be generally the mean value of squared select function difference. We can regard the DoM model as a special case of the CPA model with the number of bits as 1. In our future work, we will investigate the success rate formulas for other power analysis attacks and timing attacks and their constituent parameters.

7 Conclusions

In this paper, a theoretical model for DPA on cryptographic systems is presented. It reveals how physical implementations and cryptographic algorithms jointly affect the SCA resilience. The relation between the success rate and cryptographic systems is modeled over a multivariate Gaussian distribution. The signal-to-noise ratio between the power difference and standard deviation of the power distribution indicates how resilient an implementation is. The confusion matrix

generated by algorithmic confusion analysis illustrates how the cryptographic algorithm affects the resilience. Experimental results on DES and AES verify the model. We believe that this model is innovative, provides valuable insights on side-channel characteristics of cryptosystems, and could significantly facilitate SCA-resilient design and implementations.

References

1. Kocher, P.C., Jaffe, J., Jun, B.: Differential Power Analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
2. Brier, E., Clavier, C., Olivier, F.: Correlation Power Analysis with a Leakage Model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004)
3. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual Information Analysis. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008)
4. Le, T.-H., Clédière, J., Canovas, C., Robisson, B., Servière, C., Lacoume, J.-L.: A Proposition for Correlation Power Analysis Enhancement. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 174–186. Springer, Heidelberg (2006)
5. Quisquater, J.-J., Samyde, D.: ElectroMagnetic Analysis (EMA): Measures and Counter-Measures for Smart Cards. In: Attali, S., Jensen, T. (eds.) E-smart 2001. LNCS, vol. 2140, pp. 200–210. Springer, Heidelberg (2001)
6. Gandolfi, K., Mourtel, C., Olivier, F.: Electromagnetic Analysis: Concrete Results. In: Koç, Ç.K., Naccache, D., Paar, C. (eds.) CHES 2001. LNCS, vol. 2162, pp. 251–261. Springer, Heidelberg (2001)
7. Kocher, P.C.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996)
8. Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards Sound Approaches to Counteract Power-Analysis Attacks. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 398–412. Springer, Heidelberg (1999)
9. Tiri, K., Verbauwhede, I.: A VLSI design flow for secure side-channel attack resistant ICs. In: Proc. Design, Automation & Test in Europe, pp. 58–63 (2005)
10. Clavier, C., Coron, J.-S., Dabbous, N.: Differential Power Analysis in the Presence of Hardware Countermeasures. In: Paar, C., Koç, Ç.K. (eds.) CHES 2000. LNCS, vol. 1965, pp. 252–263. Springer, Heidelberg (2000)
11. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. Stochastic Methods: A Performance Analysis for Side Channel Cryptanalysis. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 15–29. Springer, Heidelberg (2006)
12. Standaert, F.-X., Bulens, P., de Meulenaer, G., Veyrat-Charvillon, N.: Improving the rules of the DPA contest. Cryptology ePrint Archive, Report 2008/517 (2008), http://eprint.iacr.org/2008/517
13. Standaert, F.-X., Malkin, T.G., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 443–461. Springer, Heidelberg (2009)
14. Veyrat-Charvillon, N., Standaert, F.-X.: Mutual Information Analysis: How, When and Why? In: Clavier, C., Gaj, K. (eds.) CHES 2009. LNCS, vol. 5747, pp. 429–443. Springer, Heidelberg (2009)

15. Messerges, T.S., Dabbish, E.A., Sloan, R.H.: Examining smart-card security under the threat of power analysis attacks. IEEE Trans. on Computers 51(5), 541–552 (2002)
16. Bevan, R., Knudsen, E.: Ways to Enhance Differential Power Analysis. In: Lee, P.J., Lim, C.H. (eds.) ICISC 2002. LNCS, vol. 2587, pp. 327–342. Springer, Heidelberg (2003)
17. Mangard, S.: Hardware Countermeasures against DPA – A Statistical Analysis of Their Effectiveness. In: Okamoto, T. (ed.) CT-RSA 2004. LNCS, vol. 2964, pp. 222–235. Springer, Heidelberg (2004)
18. Prouff, E.: DPA Attacks and S-Boxes. In: Gilbert, H., Handschuh, H. (eds.) FSE 2005. LNCS, vol. 3557, pp. 424–441. Springer, Heidelberg (2005)
19. Shannon, E.A.: Communication theory of secrecy systems. Bell System Technical Journal 28(4), 656–715 (1949)
20. Katz, J., Lindell, Y.: Introduction to Modern Cryptography. Chapman & Hall/CRC Press (2007)
21. Johnson, O.T.: Information Theory and the Central Limit Theorem. Imperial College Press (2004)
22. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards. Advances in Information Security. Springer-Verlag New York (2007)
23. DPA Contest, `http://www.dpacontest.org/`
24. Side-channel attack standard evaluation board (SASEBO). Research Center for Information Security (RCIS), `http://www.rcis.aist.go.jp/special/SASEBO/index-en.html`
25. Coron, J.-S., Kizhvatov, I.: An Efficient Method for Random Delay Generation in Embedded Software. In: Clavier, C., Gaj, K. (eds.) CHES 2009. LNCS, vol. 5747, pp. 156–170. Springer, Heidelberg (2009)
26. Coron, J.-S., Kizhvatov, I.: Analysis and Improvement of the Random Delay Countermeasure of CHES 2009. In: Mangard, S., Standaert, F.-X. (eds.) CHES 2010. LNCS, vol. 6225, pp. 95–109. Springer, Heidelberg (2010)
27. Mangard, S., Oswald, E., Standaert, F.-X.: One for all - all for one: unifying standard differential power analysis attacks. IET Inf. Security 5(2), 100–110 (2011)
28. Standaert, F.-X., Peeters, E., Rouvroy, G., Quisquater, J.: An overview of power analysis attacks against field programmable gate arrays. Proc. IEEE 94(2) (2006)

Appendix

A The Related CPA Model

A model for CPA is proposed in [17] and improved in [28]. The overall success rate of CPA is given as:

$$\mathrm{SR} = \left(\int_0^{\infty} \frac{1}{\frac{1}{\sqrt{N_m - 3}}\sqrt{2\pi}} \exp\left\{ -\frac{(x-r)^2}{\frac{2}{N_m-3}} \right\} \mathrm{d}x \right)^{N_k - 1}$$

where r is the Pearson correlation of CPA for the correct key, N_k is the number of key guesses in CPA, and N_m is the number of measurements.

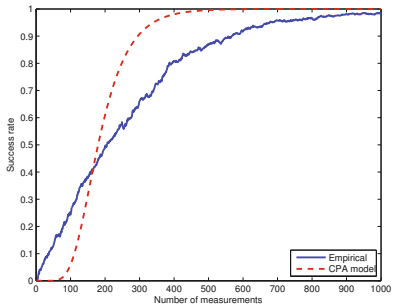

Fig. 3. Empirical and theoretical success rates of CPA on DES

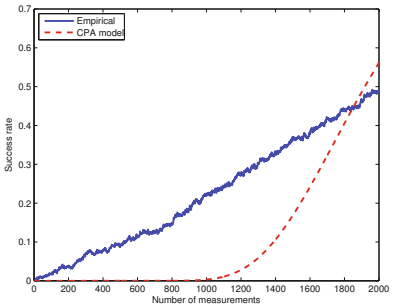

Fig. 4. Empirical and theoretical success rates of CPA on AES

The CPA model assumes that the different dimensions of the overall success rate are independent, i.e., the covariances between different key guesses are 0. This is a false assumption, and therefore makes the CPA model inaccurate. We generate success rates using the CPA model for DES and AES with the same data set used in Section 5, as shown in Figs.3 and 4. The success rate curves from the CPA model do not match the empirical results.

B Proof for the Confusion Lemma (Lemma 1)

Apply Theorem 1 to each of the key hypotheses k_i and k_j, we have:

$$\Pr[\psi|k_i = 1] = \Pr[\psi|k_i = 0] = \frac{1}{2}, \qquad \Pr[\psi|k_j = 1] = \Pr[\psi|k_j = 0] = \frac{1}{2}.$$

Because $\Pr[\psi|k_i = 1] = \Pr[\psi|k_i = 1, \psi|k_j = 0] + \Pr[\psi|k_i = 1, \psi|k_j = 1]$, and similarly for the other three probabilities above, from the definitions of the coefficients κ and ξ, we have:

$$\Pr[\psi|k_i = 1, \psi|k_j = 0] = \Pr[\psi|k_i = 0, \psi|k_j = 1] = \frac{1}{2}\kappa,$$
$$\Pr[\psi|k_i = 0, \psi|k_j = 0] = \Pr[\psi|k_i = 1, \psi|k_j = 1] = \frac{1}{2}\xi.$$

C Proof for Lemma 2

$$\begin{aligned}
\kappa(k_h, k_i) + \kappa(k_h, k_j) &= \Pr[(\psi|k_h) \neq (\psi|k_i)] + \Pr[(\psi|k_h) \neq (\psi|k_j)] \\
&= \Pr[(\psi|k_j) = (\psi|k_h) \neq (\psi|k_i)] + \Pr[(\psi|k_j) = (\psi|k_i) \neq (\psi|k_h)] + \\
&\quad \Pr[(\psi|k_i) = (\psi|k_h) \neq (\psi|k_j)] + \Pr[(\psi|k_i) = (\psi|k_j) \neq (\psi|k_h)] \\
&= \Pr[(\psi|k_j) \neq (\psi|k_i)] + 2\Pr[(\psi|k_j) = (\psi|k_i) \neq (\psi|k_h)] \\
&= \kappa(k_i, k_j) + 2\tilde{\kappa}(k_h, k_i, k_j).
\end{aligned}$$

Therefore: $\tilde{\kappa}(k_h, k_i, k_j) = \frac{1}{2}[\kappa(k_h, k_i) + \kappa(k_h, k_j) - \kappa(k_i, k_j)]$.

D Confusion Coefficient and DPA DoM

In DPA, the set of waveforms $\mathcal{W}$ is divided into two groups according to the value of ψ for one key hypothesis. Therefore, for the correct key k_c and a guessed key k_g the N_m measurements are divided into four groups $\mathcal{W}_{ij}$, $i, j = 0, 1$, as shown in Table 1. For example, $\mathcal{W}_{10}$ is the group of measurements that satisfy $\psi|k_c = 1$ and $\psi|k_g = 0$, and $\mathcal{W}_{1\_}$ is the group of measurements for $\psi|k_c = 1$. We denote the number of measurements in each group by N_{ij}. We have $N_{1\_} = N_{11} + N_{10}, N_{0\_} = N_{01} + N_{00}, N_{1\_} + N_{0\_} = N_{\_1} + N_{\_0} = N_m$. Suppose the DoM for k_c and k_g are δ_c

Table 1. The four groups of waveforms $\mathcal{W}_{ij}$ and their number of measurements

	$\psi\|k_c = 1$	$\psi\|k_c = 0$	total
$\psi\|k_g = 1$	$\mathcal{W}_{11}$ (N_{11})	$\mathcal{W}_{01}$ (N_{01})	$\mathcal{W}_{\_1}$ $(N_{\_1})$
$\psi\|k_g = 0$	$\mathcal{W}_{10}$ (N_{10})	$\mathcal{W}_{00}$ (N_{00})	$\mathcal{W}_{\_0}$ $(N_{\_0})$
total	$\mathcal{W}_{1\_}$ $(N_{1\_})$	$\mathcal{W}_{0\_}$ $(N_{0\_})$	$\mathcal{W}$ (N_m)

and δ_g, respectively. The difference between the two DoMs is:

$$\begin{aligned}\Delta(k_c, k_g) = \delta_c - \delta_g =& (\frac{N_{11}}{N_{1\_}} - \frac{N_{11}}{N_{\_1}})\bar{\mathcal{W}}_{11} + (\frac{N_{10}}{N_{1\_}} + \frac{N_{10}}{N_{\_0}})\bar{\mathcal{W}}_{10} \\ & - (\frac{N_{01}}{N_{0\_}} + \frac{N_{01}}{N_{\_1}})\bar{\mathcal{W}}_{01} - (\frac{N_{00}}{N_{0\_}} - \frac{N_{00}}{N_{\_0}})\bar{\mathcal{W}}_{00} \end{aligned} \quad (8)$$

where $\bar{\mathcal{W}}_{ij} = \sum \mathcal{W}_{ij}/N_{ij}$ for $i, j = 0, 1$, which are normal random variables according to the Central Limit Theorem as given in Section 2.3. Hence $\Delta(k_c, k_g)$ is also normally distributed because it is a linear combination of normal random variables. We now calculate its mean and variance.

Note that there are two sources of randomness in $\Delta(k_c, k_g)$. The first source is from the randomly selected plaintexts. Denote $\boldsymbol{\psi}_c = (\psi_1, ..., \psi_{N_m})|k_c$ and $\boldsymbol{\psi}_g = (\psi_1, ..., \psi_{N_m})|k_g$ as the values of the select function for the set of measurement plaintext under the correct key k_c and incorrect key k_g, respectively. $\boldsymbol{\psi}_c$, $\boldsymbol{\psi}_g$, and N_{ij}s are all random variables. Conditional on given $\boldsymbol{\psi}_c$ and $\boldsymbol{\psi}_g$, the partition of the measured waveforms $\mathcal{W}$ into four groups are fixed, and thus N_{ij}s are constants. There is still the second source of randomness, measurement errors in $\mathcal{W}$. Therefore $\bar{\mathcal{W}}_{ij}$s are still random variables conditional on given $\boldsymbol{\psi}_c$ and $\boldsymbol{\psi}_g$.

Given $\boldsymbol{\psi}_c$ and $\boldsymbol{\psi}_g$, the waveforms in groups $\mathcal{W}_{11}$ and $\mathcal{W}_{10}$ have the same mean, which is larger by the amount ϵ than the mean of waveforms in $\mathcal{W}_{01}$ and $\mathcal{W}_{00}$. Without loss of generality, we assume that the theoretical means of $\bar{\mathcal{W}}_{11}$, $\bar{\mathcal{W}}_{10}$, $\bar{\mathcal{W}}_{01}$ and $\bar{\mathcal{W}}_{00}$ are ϵ, ϵ, 0 and 0, respectively. Therefore from equation (8), the conditional mean of $\Delta(k_c, k_g)$ is:

$$\begin{aligned} E[\Delta(k_c, k_g)|\boldsymbol{\psi}_c, \boldsymbol{\psi}_g] &= (\frac{N_{11}}{N_{1\_}} - \frac{N_{11}}{N_{\_1}} + \frac{N_{10}}{N_{1\_}} + \frac{N_{10}}{N_{\_0}})\epsilon \\ &= (1 - \frac{N_{11}}{N_{\_1}} + \frac{N_{10}}{N_{\_0}})\epsilon = (\frac{N_{01}}{N_{\_1}} + \frac{N_{10}}{N_{\_0}})\epsilon. \end{aligned}$$

Now we consider the randomness in N_{ij}s, related to the algorithmic confusion analysis. According to Lemma 1, each of the waveform with $\psi|k_g = 1$ has $\kappa(k_c, k_g)$ probability to have $\psi|k_c = 0$. N_{01} and N_{10} independently follow Binomial($N_{\_1}$,$\kappa(k_c, k_g)$) and Binomial($N_{\_0}$,$\kappa(k_c, k_g)$) distributions, respectively.

$$E(\frac{N_{01}}{N_{\_1}} + \frac{N_{10}}{N_{\_0}}) = \frac{\kappa(k_c, k_g)N_{\_1}}{N_{\_1}} + \frac{\kappa(k_c, k_g)N_{\_0}}{N_{\_0}} = 2\kappa(k_c, k_g),$$

$$Var(\frac{N_{01}}{N_{\_1}} + \frac{N_{10}}{N_{\_0}}) = \frac{\kappa(k_c, k_g)\xi(k_c, k_g)}{N_{\_1}} + \frac{\kappa(k_c, k_g)\xi(k_c, k_g)}{N_{\_0}}$$

As $N_m \to \infty$, according to Theorem 1 and Lemma 1, $N_{1\_} \simeq N_{0\_} \simeq N_{\_1} \simeq N_{\_0} \simeq N_m/2$, $N_{10} \simeq N_{01} \simeq \kappa(k_c, k_g)N_m/2$, $N_{11} \simeq N_{00} \simeq [1 - \kappa(k_c, k_g)]N_m/2$. Thus:

$$E[\Delta(k_c, k_g)] = E\{E[\Delta(k_c, k_g)|\boldsymbol{\psi}_c, \boldsymbol{\psi}_g]\} = 2\kappa(k_c, k_g)\epsilon \tag{9}$$

$$Var\{E[\Delta(k_c, k_g)|\boldsymbol{\psi}_c, \boldsymbol{\psi}_g]\} = 4\kappa(k_c, k_g)\xi(k_c, k_g)\frac{\epsilon^2}{N_m} \tag{10}$$

The $\bar{\mathcal{W}}_{ij}$s are independent and each has the conditional variance $\sigma_{\mathcal{W}}/\sqrt{N_{ij}}$. From (8), we get:

$$\begin{aligned} E\{Var[\Delta(k_c, k_g)|\boldsymbol{\psi}_c, \boldsymbol{\psi}_g]\} =& E\{\sigma^2_{\mathcal{W}}[(\frac{1}{N_{1\_}} - \frac{1}{N_{\_1}})^2 N_{11} + (\frac{1}{N_{1\_}} + \frac{1}{N_{\_0}})^2 N_{10} \\ &+ (\frac{1}{N_{0\_}} + \frac{1}{N_{\_1}})^2 N_{01} + (\frac{1}{N_{0\_}} - \frac{1}{N_{\_0}})^2 N_{00}]\} \\ =& \sigma^2_{\mathcal{W}} \frac{16\kappa(k_c, k_g)}{N_m}. \end{aligned}$$

Combined with (10),

$$\begin{aligned} Var[\Delta(k_c, k_g)] =& E\{Var[\Delta(k_c, k_g)|\boldsymbol{\psi}_c, \boldsymbol{\psi}_g]\} + Var\{E[\Delta(k_c, k_g)|\boldsymbol{\psi}_c, \boldsymbol{\psi}_g]\} \\ =& 16\kappa(k_c, k_g)\frac{\sigma^2_{\mathcal{W}}}{N_m} + 4\kappa(k_c, k_g)\xi(k_c, k_g)\frac{\epsilon^2}{N_m}. \end{aligned} \tag{11}$$

E 2-keys Success Rate

For the 2-keys success rate, we have got $Y_2 = [y_1, y_2]^T$, a random vector with two-dimension normal distribution, $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. Now we calculate the formula for $\boldsymbol{\Sigma}_2$. From equation (8), we have (for $i = 1, 2$):

$$\begin{aligned} y_i = \Delta(k_c, k_{gi}) =& (\frac{N_{11,y_i}}{N_{1\_,y_i}} - \frac{N_{11,y_i}}{N_{\_1,y_i}})\bar{\mathcal{W}}_{11,y_i} + (\frac{N_{10,y_i}}{N_{1\_,y_i}} + \frac{N_{10,y_i}}{N_{\_0,y_i}})\bar{\mathcal{W}}_{10,y_i} \\ &- (\frac{N_{00,y_i}}{N_{0\_,y_i}} - \frac{N_{00,y_i}}{N_{\_0,y_i}})\bar{\mathcal{W}}_{00,y_i} - (\frac{N_{01,y_i}}{N_{\_1,y_i}} + \frac{N_{01,y_i}}{N_{0\_,y_i}})\bar{\mathcal{W}}_{01,y_i} \end{aligned} \tag{12}$$

From Appendix D Equation (11), the variance of y_1 and y_2 are:

$$\mathrm{Cov}(y_i, y_i) = 16\kappa(k_c, k_{g_i})\frac{\sigma^2_{\mathcal{W}}}{N_m} + 4\kappa(k_c, k_{g_i})\xi(k_c, k_{g_i})\frac{\epsilon^2}{N_m}, \qquad i = 1, 2$$

Next we compute $\mathrm{Cov}(y_1, y_2)$, the covariance between y_1 and y_2. We first calculate $E[Cov(y_1, y_2|\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2})]$. The conditional covariance between y_1 and y_2 given $(\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2})$ is:

$$\begin{aligned}
&Cov(y_1, y_2|\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2}) \\
=&4\kappa(k_c, k_{g_1})\kappa(k_c, k_{g_2})[Cov(\bar{\mathcal{W}}_{10,y_1}, \bar{\mathcal{W}}_{10,y_2}) + Cov(\bar{\mathcal{W}}_{01,y_1}, \bar{\mathcal{W}}_{01,y_2}) \\
&- Cov(\bar{\mathcal{W}}_{10,y_1}, \bar{\mathcal{W}}_{01,y_2}) - Cov(\bar{\mathcal{W}}_{01,y_1}, \bar{\mathcal{W}}_{10,y_2})]
\end{aligned}$$

The waveforms in $\mathcal{W}_{10_{y_1}}$ and $\mathcal{W}_{10_{y_2}}$ are those with $\psi|k_c = 1$, different from those in $\mathcal{W}_{01_{y_1}}$ and $\mathcal{W}_{01_{y_2}}$, therefore: $Cov(\bar{\mathcal{W}}_{10,y_1}, \bar{\mathcal{W}}_{01,y_2}) = Cov(\bar{\mathcal{W}}_{01,y_1}, \bar{\mathcal{W}}_{10,y_2}) = 0$.

To compute $Cov(\bar{\mathcal{W}}_{10,y_1}, \bar{\mathcal{W}}_{10,y_2})$, we consider how similar they are, i.e., how many waveforms are the same between the partitions $\mathcal{W}_{10_{y_1}}$ and $\mathcal{W}_{10_{y_2}}$. Let $N_{10,s}$ denote the number of same waveforms between $\mathcal{W}_{10,y_1}$ and $\mathcal{W}_{10,y_2}$. Then $Cov(\sum \mathcal{W}_{10,y_1}, \sum \mathcal{W}_{10,y_1}) = N_{10,s}\sigma^2_{\mathcal{W}}$. $N_{10,s} \simeq \tilde{\kappa}(k_c, k_{g_1}, k_{g_2})N_m/2$ as $N_m \to \infty$ by the definition of $\tilde{\kappa}(k_c, k_{g_1}, k_{g_2})$. Hence,

$$Cov(\bar{\mathcal{W}}_{10,y_1}, \bar{\mathcal{W}}_{10,y_2}) = \frac{Cov(\sum \mathcal{W}_{10,y_1}, \sum \mathcal{W}_{10,y_2})}{N_{10,y_1}N_{10,y_2}} = \frac{2\tilde{\kappa}(k_c, k_{g_1}, k_{g_2})}{\kappa(k_c, k_{g_1})\kappa(k_c, k_{g_2})N_m}\sigma^2_{\mathcal{W}}$$

Similarly, we get the same expression for $Cov(\bar{\mathcal{W}}_{01,y_1}, \bar{\mathcal{W}}_{01,y_2})$. Thus we get

$$E[Cov(y_1, y_2|\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2})] = 16\tilde{\kappa}(k_c, k_{g_1}, k_{g_2})\frac{\sigma^2_{\mathcal{W}}}{N_m}. \tag{13}$$

Next we calculate $Cov[E(y_1|\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2}), E(y_2|\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2})]$. Let N_{d,y_1} denote the number of measurements where $\psi|k_c$ is different from $\psi|k_{g_1}$. From (12),

$$\begin{aligned}
&Cov[E(y_1|\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2}), E(y_2|\boldsymbol{\psi}_c, \boldsymbol{\psi}_{g_1}, \boldsymbol{\psi}_{g_2})] \\
=&(\frac{2\epsilon}{N_m})^2 Cov(N_{d,y_1}, N_{d,y_2}) = (\frac{2\epsilon}{N_m})^2[E(N_{d,y_1}N_{d,y_2}) - E(N_{d,y_1})E(N_{d,y_2})]
\end{aligned}$$

We re-express $N_{d,y_1} = \sum I[(\psi|k_c) \neq (\psi|k_{g_1})]$ and $N_{d,y_2} = \sum I[(\psi|k_c) \neq (\psi|k_{g_2})]$. Obviously $E(N_{d,y_1}) = N_m\kappa(k_c, k_{g_1})$ and $E(N_{d,y_2}) = N_m\kappa(k_c, k_{g_2})$.

$$N_{d,y_1}N_{d,y_2} = \sum I[(\psi|k_c) \neq (\psi|k_{g_1})] \sum I[(\psi|k_c) \neq (\psi|k_{g_2})]$$

Note that $N_{d,y_1}N_{d,y_2}$ is the sum of N_m^2 terms. Most of the terms in the sum have expectation $\kappa(k_c, k_{g_1})\kappa(k_c, k_{g_2})$ except for those N_m terms corresponding to the same waveforms, which have expectation:

$$\begin{aligned}
E\{I[(\psi|k_c) \neq (\psi|k_{g_1})]I[(\psi|k_c) \neq (\psi|k_{g_2})]\} &= E\{I[(\psi|k_{g_1}) = (\psi|k_{g_2}) \neq (\psi|k_c)]\} \\
&= \tilde{\kappa}(k_c, k_{g_1}, k_{g_2})
\end{aligned}$$

Hence,

$$E(N_{d,y_1}N_{d,y_2}) = N_m^2\kappa(k_c,k_{g_1})\kappa(k_c,k_{g_2}) + N_m[\tilde{\kappa}(k_c,k_{g_1},k_{g_2}) - \kappa(k_c,k_{g_1})\kappa(k_c,k_{g_2})]$$

This implies that

$$\begin{aligned} &Cov[E(y_1|\boldsymbol{\psi}_c,\boldsymbol{\psi}_{g_1},\boldsymbol{\psi}_{g_2}), E(y_2|\boldsymbol{\psi}_c,\boldsymbol{\psi}_{g_1},\boldsymbol{\psi}_{g_2})] \\ &= 4[\tilde{\kappa}(k_c,k_{g_1},k_{g_2}) - \kappa(k_c,k_{g_1})\kappa(k_c,k_{g_2})]\frac{\epsilon^2}{N_m}. \end{aligned} \tag{14}$$

Combining (13) and (14), we get:

$$\begin{aligned} &Cov(y_1,y_2) \\ =&E[Cov(y_1,y_2|\boldsymbol{\psi}_c,\boldsymbol{\psi}_{g_1},\boldsymbol{\psi}_{g_2})] + Cov[E(y_1|\boldsymbol{\psi}_c,\boldsymbol{\psi}_{g_1},\boldsymbol{\psi}_{g_2}), E(y_2|\boldsymbol{\psi}_c,\boldsymbol{\psi}_{g_1},\boldsymbol{\psi}_{g_2})] \\ =&16\tilde{\kappa}(k_c,k_{g_1},k_{g_2})\frac{\sigma_{\mathcal{W}}^2}{N_m} + 4[\tilde{\kappa}(k_c,k_{g_1},k_{g_2}) - \kappa(k_c,k_{g_1})\kappa(k_c,k_{g_2})]\frac{\epsilon^2}{N_m}. \end{aligned}$$

F Leakage Evaluation of Random Delay

The resilience of random delay is determined by the maximum leakage ϵ_{rd}, which is the overall leakage accumulated with random shifting. We consider two scenarios of the original leakage:

1. Single-point leakage. Only one time point s in the power consumption waveform leaks information with signal level ϵ. This is the simplified ideal case. The maximum leakage after random shifting is the original leakage distributed with the maximum probability, which is:

   $$\epsilon_{rd} = \epsilon \cdot f_{rd}(t_{max}) = \epsilon \cdot \max_{0 \le t \le N_{rd}-1} \{f_{rd}(t)\}.$$

   For uniform random delay, $\Pr(t) = f_{rd}(t) = 1/N_{rd}$, for $t = 0, 1, ..., N_{rd} - 1$. Hence the signal $\epsilon_{rd} = \epsilon/N_{rd}$ decreases from the original signal ϵ by a factor N_{rd}.
2. Multiple-point leakage. At time t, the leakage signal strength is $\epsilon(t)$. Then the leakage signal at time i with random delay is:

   $$\epsilon_{rd}(i) = \sum_{t=0}^{N_{rd}-1} f_{rd}(t)\epsilon(i+t).$$

   The maximum leakage accumulation with random delay as:

   $$\epsilon_{rd} = \max_i \left\{ \sum_{t=0}^{N_{rd}-1} f_{rd}(t)\epsilon(i+t) \right\}.$$

Then the success rate of the strongest single-point DoM attack on the device with random delay can be calculated by Formula (7) using the ϵ_{rd} value.