Li *et al. Cybersecur* (2021) 4:35
https://doi.org/10.1186/s42400-021-00099-1

Cybersecurity

RESEARCH Open Access

Transparency order versus confusion coefficient: a case study of NIST lightweight cryptography S-Boxes

Huizhong Li[1,2], Guang Yang[4], Jingdian Ming[1,2], Yongbin Zhou[1,2,3*] and Chengbin Jin[1,2]

Abstract

Side-channel resistance is nowadays widely accepted as a crucial factor in deciding the security assurance level of cryptographic implementations. In most cases, non-linear components (e.g. S-Boxes) of cryptographic algorithms will be chosen as primary targets of side-channel attacks (SCAs). In order to measure side-channel resistance of S-Boxes, three theoretical metrics are proposed and they are reVisited transparency order (VTO), confusion coefficients variance (CCV), and minimum confusion coefficient (MCC), respectively. However, the practical effectiveness of these metrics remains still unclear. Taking the 4-bit and 8-bit S-Boxes used in NIST Lightweight Cryptography candidates as concrete examples, this paper takes a comprehensive study of the applicability of these metrics. First of all, we empirically investigate the relations among three metrics for targeted S-boxes, and find that CCV is almost linearly correlated with VTO, while MCC is inconsistent with the other two. Furthermore, in order to verify which metric is more effective in which scenarios, we perform simulated and practical experiments on nine 4-bit S-Boxes under the non-profiled attacks and profiled attacks, respectively. The experiments show that for quantifying side-channel resistance of S-Boxes under non-profiled attacks, VTO and CCV are more reliable while MCC fails. We also obtain an interesting observation that none of these three metrics is suitable for measuring the resistance of S-Boxes against profiled SCAs. Finally, we try to verify whether these metrics can be applied to compare the resistance of S-Boxes with different sizes. Unfortunately, all of them are invalid in this scenario.

Keywords: Side-channel attacks, NIST lightweight cryptography S-Boxes, Transparency order, Confusion coefficient

Introduction

With the emergence and explosive development of the Internet of Things, a large number of highly constrained devices are interconnected and working in concert to accomplish certain tasks (Zhu and Reddi 2017). In order to protect the security of most applications, lightweight cryptographic algorithms tailored for constrained devices have been researched for more than a decade (Heuser et al. 2020). Specifically, NIST has initiated a process to solicit, evaluate, and standardize lightweight cryptographic algorithms (NIST 2021). Subsequently, many ingenious ciphers have been proposed (Bao et al. 2019; Zhang et al. 2019; Dobraunig and Mennink 2019).

The security evaluation of lightweight cryptographic algorithms is a topic of interest due to their wide application prospects. In particular, the resistance of cryptographic implementations against side-channel attacks (SCAs) has been recognized as a crucial factor (Heuser et al. 2020). Essentially, SCAs exploit physical leakages (e.g., power consumption (Kocher et al. 1999), electromagnetic emanations (Brier et al. 2004)) from cryptosystems to recover their underlying sensitive data. Generally speaking, SCAs can be divided into two classes: non-profiled attacks, such as differential power analysis (DPA) (Kocher et al. 1999) and correlation power analysis (CPA)

*Correspondence: zhouyongbin@iie.ac.cn
[3] School of Cyber Security, Nanjing University of Science and Technology, Nanjing 210094, China
Full list of author information is available at the end of the article

(Brier et al. 2004), and profiled attacks, such as template attacks (TA) (Chari et al. 2002) and deep learning (DL) based profiled attacks (Maghrebi et al. 2016; Cagli et al. 2017; Wouters et al. 2020).

When performing an efficient SCA, it is evident that non-linear components (e.g. S-Boxes) of cryptographic algorithms will be chosen as the primary targets (Carlet 2005). Therefore, for evaluating the side-channel resistance of a lightweight cipher, it is an important perspective to study how to measure the intrinsic resistance of S-Boxes against SCAs. Consequently, various metrics have been proposed, such as DPA signal-to-noise ratio (Guilley et al. 2004), transparency orders (Prouff 2005; Chakraborty et al. 2017; Li et al. 2020), confusion coefficients (Fei et al. 2012) and non-absolute indicator (Carlet et al. 2021).

Among those metrics, transparency orders and confusion coefficients are the most commonly used to compare and select optimal S-Boxes with high SCA resistance. As for the first ones, the original transparency order (TO) (Prouff 2005) and modified transparency order (MTO) (Chakraborty et al. 2017) has been widely used to select 4×4 S-Boxes, 6×6 S-Boxes, and 8×8 S-Boxes (Picek et al. 2014, 2016; Kavut and Baloğlu 2016; Patranabis et al. 2019). However, it has been pointed out that both TO and MTO are flawed (Li et al. 2020). And the notion of reVisited transparency order (VTO) was further proposed in Li et al. (2020). As far as we know, VTO has been used to select 4×4 S-Boxes in Runlian et al. (2020) and 8×8 S-Boxes in Martínez-Díaz and Freyre-Echevarria (2020). As for confusion coefficients, confusion coefficient variance (CCV) and minimum confusion coefficient (MCC) were proposed by Picek et al. (2014) and Guilley et al. (2015), respectively. CCV has been used to heuristically select optimal 4×4 and 8×8 S-Boxes for cryptographic algorithms (Ege et al. 2015; Freyre-Echevarría et al. 2020). While MCC has not received much attention. Furthermore, there are some studies consider both transparency orders and confusion coefficients to select optimal S-Boxes against SCAs (de la Cruz Jiménez 2018; Martínez-Díaz and Freyre-Echevarria 2020).

However, the practical effectiveness of these metrics remains still unclear. Specifically, for transparency orders, the existing research work is limited to the analysis of TO or MTO, and there is a lack of research on the recently proposed VTO. And for confusion coefficients, the effectiveness of CCV and MCC needs to be further verified. Therefore, we mainly focus on investigating the applicability and relations of VTO, CCV, and MCC in this work.

Our Contributions. In this paper, we give a comprehensive study of the applicability of three typical theoretical metrics for side-channel analysis, namely VTO, CCV and MCC. We take the 4-bit and 8-bit S-Boxes used in NIST Lightweight Cryptography candidates as concrete examples for our analysis. Firstly, we empirically investigate the relations among three metrics for targeted S-boxes. The metric values of these S-Boxes show that CCV is almost linearly correlated with VTO, while MCC is inconsistent with the other two metrics.

Next, to verify the effectiveness of these metrics, we perform simulated and practical experiments on nine 4-bit S-Boxes in the non-profiled and profiled scenarios, respectively. For the non-profiled scenario, when VTO (resp. CCV) difference value of two S-Boxes is relatively large, the S-Box with a lower VTO (resp. higher CCV) value is generally more resistant to attacks. However, when VTO and CCV values of S-Boxes turn relatively close to each other, these two metrics become inaccurate to some extent. Interestingly, the MCC fails to work in quantifying the resistance of S-Boxes against CPA attacks. For the profiled scenario, template attacks and deep learning based profiled attacks are performed, respectively. Unfortunately, none of these three metrics (VTO, CCV and MCC) is suitable for measuring the resistance of S-Boxes against profiled SCAs.

Finally, we try to verify whether these metrics can be applied to compare the resistance of S-Boxes with different sizes. Interestingly, all of them cannot be used to compare the resistance of S-Boxes with different sizes.

The rest of the paper is organized as follows. "Notations and preliminaries" section gives preliminary notions on S-Boxes and theoretical metrics evaluating the resiliency of S-Boxes against SCAs. "Evaluation of S-Boxes" section provides basic information on the S-Boxes we evaluated and the results based on the theoretical metrics. Then in Non-profiled side-channel attacks against 4×4 S-Boxes section, we demonstrate the simulated and practical results of non-profiled attacks on nine 4-bit S-Boxes. And the results of profiled attacks are shown in Profiled side-channel attacks section. Furthermore, we verify whether these metrics can be applied to compare the resistance of S-Boxes with different sizes in "p0$4 \times 4$ S-Boxes versus 8×8 S-Boxes" section. Finally, we conclude our work in "Conclusions and future work" section.

Notations and preliminaries

In this section, we first give basic notions about the cryptographic properties of S-Boxes. Then, we introduce the notions of reVisited transparency order (VTO), confusion coefficient variance (CCV), and minimum confusion coefficient (MCC).

Boolean functions and S-Boxes

Let $\mathbb{F}_2^n$ be the vector space that contains all the n-bit binary vectors, where n is a positive integer. For every

vector $u \in \mathbb{F}_2^n$, we denote by $\mathtt{H}(u)$ the Hamming weight (HW) of u. A Boolean function on n variables can be viewed as a mapping from $\mathbb{F}_2^n$ to $\mathbb{F}_2$, and the mappings from the vector space $\mathbb{F}_2^n$ to the vector space $\mathbb{F}_2^m$ are called (n, m)-vectorial Boolean functions where $m \leqslant n$. An (n, m)-function F that performs substitution in the cryptosystem is commonly referred to as the $n \times m$ S-Box. Generally, S-Boxes have to be chosen carefully to satisfy cryptographic properties like resisting linear and differential cryptanalysis.

For each (n, m)-function F, the Boolean functions $f_1, \ldots, f_m$ defined for every $x \in \mathbb{F}_2^n$ by $F(x) = (f_1(x), \ldots, f_m(x))$ are called the coordinate functions of F. Let $z \in \mathbb{F}_2^m$ be a vector whose binary coordinates are all zero except one which is assumed to be at index j. The j-th component function of the function F is a single output Boolean function $z \cdot F$, and we also denote this component function as F_j. The *cross-correlation spectrum* between two Boolean functions f_1, f_2 is defined as the value $\mathcal{C}_{f_1 f_2}(u) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f_1(x) \oplus f_2(x \oplus u)}$ for every $u \in \mathbb{F}_2^n$.

ReVisited transparency order

Following the work of Prouff on transparency order (TO) (Prouff 2005), Chakraborty et al. (2017) presented modified transparency order (MTO). Recently, Li et al. amended a definitional flaw in the work of TO and spotted MTO overestimates the side-channel resistance of S-Boxes in the HW leakage model. Then they proposed reVisited transparency order (VTO) and verified the soundness of this notion through simulated and practical experiments. The work of Martínez-Díaz and Freyre-Echevarria (2020) also verified that VTO is a more accurate metric. Mathematically, the VTO value of an S-Box F equals to

$$\mathrm{VTO(F)} = \max_{\beta \in \mathbb{F}_2^{\mathrm{m}}} \left(\mathrm{m} - \frac{1}{2^{2\mathrm{n}} - 2^{\mathrm{n}}} \sum_{\mathrm{a} \in \mathbb{F}_2^{\mathrm{n}^*}} \left| \sum_{\mathrm{j}=1}^{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{m}} (-1)^{\beta_\mathrm{i} \oplus \beta_\mathrm{j}} \mathcal{C}_{\mathrm{F_i,F_j}}(\mathrm{a}) \right| \right), \tag{1}$$

where β_i denotes the value of the i-th bit of the register initial state β, and $\mathcal{C}_{F_i,F_j}(a)$ denotes the cross-correlation spectrum between the component functions F_i and F_j.

Specifically, the VTO metric assumes that target devices leak the HW value of $v \oplus \beta$, where v denotes the data being processed, and β denotes the register initial state that is assumed to be constant. In Eq. (1), the value of VTO(F) is obtained by traversing all register initial state $\beta \in \mathbb{F}_2^m$, and it represents the worst case context when implementing the S-Box. However, in practice, the strategy of the adversary depends on the target device. As a result, we set the value of β to zero for each S-Box implementation in our experiments. It corresponds to our context in which the target microcontroller leaks the HW value of the manipulated value v. And the corresponding value of VTO is denoted as $\mathrm{VTO}_0(\mathrm{F})$.

Confusion coefficient variance

Fei et al. (2012) introduced another metric called confusion coefficient. This metric measurers the probability of occurrences for which key hypotheses k_i and k_i result in different intermediate values v. For DPA attacks, it can be calculated through measuring the difference between the v values under the two keys by the expectation of their squared distance. That is, it can be computed as:

$$\kappa(k_i, k_j) = \mathbb{E}\left[\left(\mathcal{L}(F(k_i \oplus p)) - \mathcal{L}\left(F\left(k_j \oplus p\right)\right) \right)^2 \right],$$

where $\mathcal{L}$ denotes the leakage function, p denotes the arbitrary inputs, and $\mathbb{E}$ is the mean operator.

Then, Picek et al. (2014) proposed to calculate the variance of all confusion coefficients with respect to each possible k_i and k_j under the HW leakage model. And the S-Box with higher confusion coefficient variance (CCV) value leads to a higher resistance against SCAs. Formally, for all the key pairs k_i, k_j, $k_i \neq k_j$, the value of CCV of an S-Box is calculated as follows:

$$\mathrm{CCV}(F) = \mathrm{Var}\left(\mathbb{E}\left[\left(\mathsf{H}(F(k_i \oplus p)) - \mathsf{H}\left(F\left(k_j \oplus p\right)\right) \right)^2 \right] \right). \tag{2}$$

Minimum confusion coefficient

Guilley et al. (2015) pointed out that when the signal-to-noise-ratio (SNR) of the leakage is low, the empirical success rate of DPA, CPA and the optimal distinguisher mainly depends on minimum confusion coefficient (MCC) $\min_{k \neq k^*} \kappa'(k^*, k)$. Where k^* denotes the secret key, and k denotes a key hypothesis that is not the secret key. The lower the value of MCC, the lower the success probability to extract the secret key based on leakages associated with the S-Box. Here the $\kappa'(k^*, k)$ is calculated as follows:

$$\kappa'\left(k^*, k\right) = \mathbb{E}\left\{ \left(\frac{\mathcal{L}(F(k^* \oplus p)) - \mathcal{L}(F(k \oplus p))}{2} \right)^2 \right\}, \tag{3}$$

which is slightly different from $\kappa(k^*, k)$, but it does not affect the order of the different S-Boxes. Note that the distribution of $\kappa'(k^*, k)$ is independent on the particular choice of k^* and the values are only permuted. Therefore, k^* can be set to 0 during the calculation. In Heuser et al. (2016) and Heuser et al. (2020), the effectiveness of using MCC to measure the resistance of different S-Boxes against CPA and the optimal distinguisher was validated through simulated experiments.

Evaluation of S-Boxes

In this section, we first show basic information on the S-Boxes we investigate. Next, the values of VTO_0, CCV, and MCC of these S-Boxes are given.

Investigated S-Boxes

Of the 25 NIST Lightweight Cryptography second-round candidates that use S-Boxes as the nonlinear component, 18 schemes use 4-bit or 8-bit S-Boxes. Therefore, we mainly evaluate the 4-bit and 8-bit S-Boxes in this work. More precisely, we focus on the following 11 S-Boxes.

4×4 S-Boxes of PHOTON-Beetle (Bao et al. 2019), KNOT (Zhang et al. 2019), Pyjamask (Goudarzi et al. 2019), GIFT-COFB (Banik et al. 2019), Elephant (Dobraunig and Mennink 2019), SATURNIN (Canteaut et al. 2019), ForkAE (Andreeva et al. 2019) and Spook (Bellizia et al. 2020). More specifically, the nine S-Boxes are listed in Table 1. Note that a cipher may use several different S-Boxes (e.g., SATURNIN). In addition, the above nine S-Boxes are also used in other NIST candidate ciphers. For instance, the GIFT S-Box is also used in ESTATE (ESTATE TweGIFT-128) (Chakraborti et al. 2020), HYENA (Chakraborti et al. 2019), SUNDAE-GIFT (Banik et al. 2019), LOTUS-AEAD and LOCUS-AEAD (Chakraborti et al. 2019). And ORANGE (Chakraborty and Nandi 2019) uses the same S-Box with PHOTON-Beetle.

8×8 S-Boxes of AES (FIPS PUB 2001) (used in SAE-AES Naito et al. 2019, mixFeed Chakraborty and Nandi 2019, COMET Gueron et al. 2019, SKINNY-AEAD and SKINNY-Hash Beierle et al. 2020 and ESTATE Chakraborti et al. 2020) and SKINNY-128 (Beierle et al. 2020) (used in SKINNY-AEAD and SKINNY-Hash Beierle et al. 2020, Romulus Iwata et al. 2019 and ForkAE Andreeva et al. 2019).

Results based on theoretical metrics

The theoretical measurement results of the VTO_0, CCV, and MCC for the S-Boxes are listed in Table 2. We can observe that when sorting the S-Boxes of the same size, the order of S-Boxes sorted by VTO_0 and sorted by CCV is basically the same. However, the ordering of S-Boxes sorted according to MCC is inconsistent with both VTO_0 and CCV. As for the 4×4 S-Boxes, the absolute Kendall rank correlation coefficients between the values of VTO_0 and CCV, VTO_0 and MCC, and CCV and MCC are 0.985, 0.039 and 0.040, respectively. That is to say, MCC conflicts with VTO_0 and CCV. The second observation is that larger S-Boxes lead to significantly higher values of VTO_0 and

Table 2 VTO_0, CCV and MCC metrics applied on the S-Boxes

Size	S-Box	VTO_0	CCV	MCC
4×4	PHOTON	2.5333	0.6627	0.2500
	KNOT	2.6000	0.6123	0.2500
	Pyjamask-128	2.6000	0.6123	0.1875
	GIFT	2.8667	0.4611	0.3125
	Elephant	2.9333	0.4611	0.3125
	$SATURNIN_{S0}$	3.0000	0.3602	0.2500
	$SATURNIN_{S1}$	3.0000	0.3602	0.2500
	SKINNY-64	3.0667	0.3098	0.2500
	Spook	3.0667	0.3098	0.2500
8×8	SKINNY-128	7.1088	0.3401	0.5000
	AES	7.4583	0.1113	0.8125

Table 1 Nine 4×4 S-Boxes for the second round candidates of NIST lightweight cryptography project

Algorithm	Notation	S-Box
PHOTON-Beetle (Bao et al. 2019)	PHOTON	12, 5, 6, 11, 9, 0, 10, 13, 3, 14, 15, 8, 4, 7, 1, 2
KNOT (Zhang et al. 2019)	KNOT	4, 0, 10, 7, 11, 14, 1, 13, 9, 15, 6, 8, 5, 2, 12, 3
Pyjamask (Goudarzi et al. 2019)	Pyjamask-128	2, 13, 3, 9, 7, 11, 10, 6, 14, 0, 15, 4, 8, 5, 1, 12
GIFT-COFB (Banik et al. 2019)	GIFT	1, 10, 4, 12, 6, 15, 3, 9, 2, 13, 11, 7, 5, 0, 8, 14
Elephant (Dobraunig and Mennink 2019)	Elephant	14, 13, 11, 0, 2, 1, 4, 15, 7, 10, 8, 5, 9, 12, 3, 6
SATURNIN (Canteaut et al. 2019)	$SATURNIN_{S0}$	0, 6, 14, 1, 15, 4, 7, 13, 9, 8, 12, 5, 2, 10, 3, 11
	$SATURNIN_{S1}$	0, 9, 13, 2, 15, 1, 11, 7, 6, 4, 5, 3, 8, 12, 10, 14
ForkAE (Andreeva et al. 2019)	SKINNY-64	12, 6, 9, 0, 1, 10, 2, 11, 3, 8, 5, 13, 4, 14, 7, 15
Spook (Bellizia et al. 2020)	Spook	0, 8, 1, 15, 2, 10, 7, 9, 4, 13, 5, 6, 14, 3, 11, 12

MCC, which implies S-Boxes with larger sizes are more vulnerable against SCAs. But CCV doesn't show such a result. Overall, there exist contradictions between the three metrics.

Specifically, for the 4×4 S-Boxes, we have the following order (from the most resistive S-Boxes to the least resistive) according to (1) VTO_0, (2) CCV, and (3) MCC:

- VTO_0: PHOTON, KNOT and Pyjamask-128, GIFT, Elephant, $\text{SATURNIN}_{\text{S0}}$ and $\text{SATURNIN}_{\text{S1}}$, SKINNY-64 and Spook.
- CCV: PHOTON, KNOT and Pyjamask-128, GIFT and Elephant, $\text{SATURNIN}_{\text{S0}}$ and $\text{SATURNIN}_{\text{S1}}$, SKINNY-64 and Spook.
- MCC: Pyjamask-128, PHOTON and KNOT and $\text{SATURNIN}_{\text{S0}}$ and $\text{SATURNIN}_{\text{S1}}$ and SKINNY-64 and Spook, GIFT and Elephant.

For the 8×8 S-Boxes, the results of all three metrics show that the S-Box of SKINNY-128 is more resistant against SCAs than that of AES.

Non-profiled side-channel attacks against 4×4 S-Boxes

Among various non-profiled attacks, we focus on CPA due to its simplicity and efficiency. Actually, CPA is equivalent to multi-bit DPA up to a change of the attacker leakage modeling (Doget et al. 2011). Therefore, VTO_0, CCV and MCC can all be used to measure the resistance of S-Boxes against CPA under the HW leakage model in theory. Concretely, CPA recovers the secret key by selecting the key that maximizes the Pearson correlation coefficient between the actual leakage and the estimated leakage based on the assumed secret key. That is,

$$\widehat{k^*} = \underset{k \in \mathcal{K}}{\text{argmax}} \left| \rho\left(\mathcal{L}_{k^*}, \widehat{\mathcal{L}_k}\right) \right|,$$

where $\rho(X, Y)$ denotes the Pearson correlation coefficient between X and Y. $\mathcal{L}_{k^*}$ represents the measured traces, and $\widehat{\mathcal{L}_k}$ denotes the estimated leakages.

Experiments of the unprotected S-Boxes

We first perform simulated and practical attacks against the nine unprotected 4×4 S-Boxes and compare their CPA resistance.

Simulated experiments

Experimental setup We implement S-Boxes in the same way by using look-up tables, and leakages are simulated as

$$\mathcal{L}(p \oplus k^*) = \text{zscore}\big(\mathsf{H}\big(F(p \oplus k^*)\big)\big) + \omega,$$

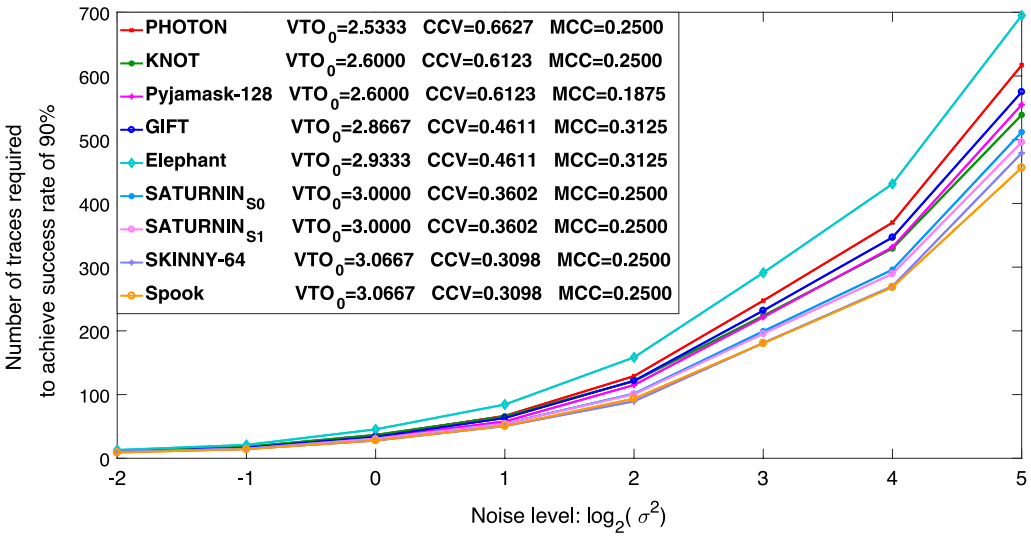
where $F(p \oplus k^*)$ denotes the sensitive variable, and ω denotes a Gaussian random variable centered in zero with a standard deviation σ. In the experimental setup, the value of σ varies in the set $\left\{2^{-1}, 2^{-\frac{1}{2}}, 1, 2^{\frac{1}{2}}, 2, 2^{\frac{3}{2}}, 4, 2^{\frac{5}{2}}\right\}$.

Experimental results In the field of side-channel analysis, success rate (Standaert et al. 2005) is a common metric to evaluate an attack. Here, for each attack, we evaluate the minimum number of traces N required to achieve an attack success rate of 90% as it is a sound way to evaluate the efficiency of a side-channel attack (Mangard 2004). The attack results are shown in Fig. 1a.

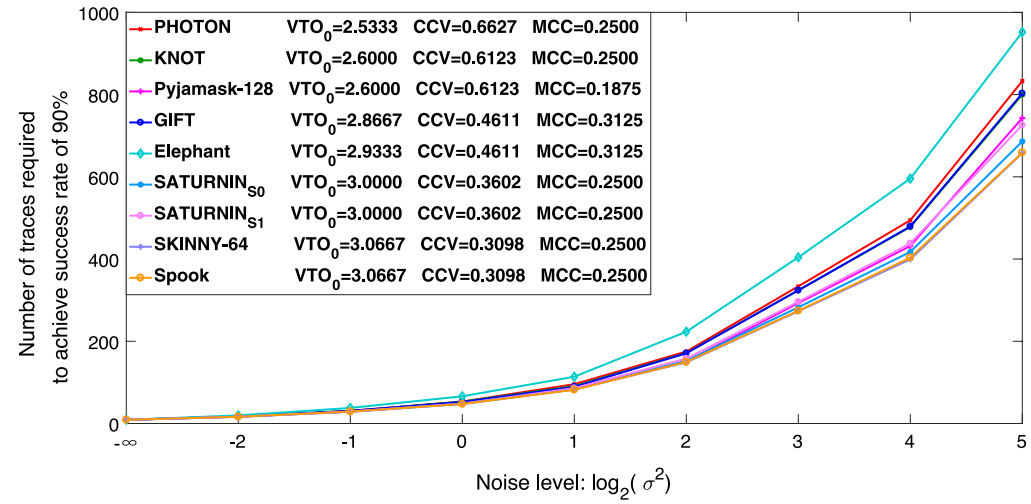
It can be observed that when the noise is low, the number of traces required for successful attacks of different S-Boxes is very close. And with the noise increases, the difference between different S-Boxes becomes more significant. However, the order of S-Boxes resistance against CPA attacks is basically the same under different noise levels. So we mainly take the result with noise variance of 2^5 as an example to illustrate for easy observation.

According to our experimental results, S-Boxes with lower VTO_0 and higher CCV values are more resistant against CPA. Such as the S-Boxes of PHOTON and GIFT are more resilient than S-Boxes of SKINNY-64 and Spook. However, the difficulty of attacking an S-Box is quite different from the outcome of the MCC metric. For example, the MCC value of Elephant's S-Box is higher than that of Spook's S-Box, while the number of required traces of the former is approximately 1.5 times that for the latter.

One may also note that sometimes there exists discordance between the VTO_0 (CCV) and the simulation results. Such as the VTO_0 (CCV) value of Elephant's S-Box is higher (lower) than that of PHOTON's S-Box, while the Elephant's S-Box is more resilient than PHOTON's S-Box. As for VTO, the reason for this phenomenon is explained in Li et al. (2020), which is due to the different perspectives of VTO and the success rate metric when quantifying the SCA resistance of S-Boxes. In detail, the basic idea of VTO is quantifying the difference between the score for the correct key and the average score for the other hypotheses; however, the success rate metric quantifies the number of successful attacks (i.e. the number of attacks in which the correct key is ranked first) in all attacks performed. As for CCV, we argue that it takes into account the distinctiveness level of the S-Box outputs for all key hypothesis pairs, which is also different from the basic idea of the success rate. Besides, the number of traces used for attacks is limited, but in the notions of VTO and CCV, it is assumed that the number of traces is sufficient so that the noise can be omitted.



(a) Simulated experiments



(b) Practical experiments

Fig. 1 CPA attacks on unprotected 4 × 4 S-Boxes

Overall, when the difference of the VTO_0 (CCV) values of the two S-Boxes is relatively large, the S-Box with a lower VTO_0 (higher CCV) value is generally more resistant to CPA attacks. However, when the difference of the VTO_0 (CCV) values of the two S-Boxes is relatively small, these two metrics lack the accuracy to evaluate the resiliency of S-Boxes. Besides, MCC fails to work in our experiments.

Practical experiments

Experimental setup In practical experiments, all the nine S-Boxes are implemented on a CW308-STM32F Target Board (for ChipWhisperer CW308 UFO Board) with the STM32F405RGT6 Arm 32-bit Cortex-M4 device, and the power traces are captured through the ChipWhisperer-Lite Capture Board (O'Flynn and Chen 2014). The sampling rate is set to 29.5 MHz, and the 500 points around the sensitive operations are taken to attack. Same as the simulated experiments, the S-Boxes are implemented by using look-up tables, and the register initial state β is set to 0. In order to study the performance of the three metrics with different noise levels, the attacks are performed based on the raw traces and traces with added Gaussian noise. Before adding noise, we standardize the traces (zero mean and unit variance). And the value of σ is set to the same as in simulated experiments.

Experimental results The attack results are shown in Fig. 1b. The $-\infty$ on the x axis represents the attack is performed on the raw traces with no additional noise. It can be observed that for most S-Box examples, the results obtained are consistent with simulated results. However, one may also note that for certain cases, the results are slightly inconsistent with the simulation results. We infer the reasons for the inconsistent results are the leakages in the real environment do not fully satisfy the HW leakage model and the noise does not fulfill the Gaussian noise assumption.

Experiments of the masked S-Boxes

Masking, due to its provable security and good device independence, has been one of the widely adopted countermeasures against SCAs (Duc et al. 2019). Naturally, the effectiveness of the three metrics is an important question when masking is adopted. Based on the work in Rivain et al. (2009), the CPA results toward dth-order masked S-Boxes of the same size is only related to the masking order d under the HW model for Boolean masking schemes. And the function of S-Box does not affect the security gain from unprotected S-Boxes to dth-order masked S-Boxes. Thus, the three metrics (VTO_0, CCV, and MCC) should be independent of the masking order for Boolean making when higher-order CPA attacks are utilized. We also try to verify it first by simulated experiments.

Simulated experiments

Experimental setup As for simulation of masking, we separately simulate first- and second-order masked S-Boxes. So two and three points corresponding to their shares y_i are simulated, and we have

$$y = \bigoplus_{i=0}^{d} y_i = SBox(p \oplus k^*), \tag{4}$$

where y denotes the output of the S-Box while $p \oplus k^*$ is the input. $y_i (i > 0)$ is generated randomly, and y_0 is processed such that Eq. (4) is satisfied. Each share of y is under the HW model, and the value of initial state β is set to zero. So each leakage point corresponding to y_i can be simulated as: $\mathcal{L}(y_i) = \mathrm{zscore}(\mathsf{H}(y_i)) + \omega_i$, where ω_i denotes the Gaussian noise centered in zero with a standard deviation σ at this moment. In the first-order masking experiments, the value of σ varies in the set $\left\{2^{-\frac{3}{4}}, 2^{-\frac{1}{2}}, 2^{-\frac{1}{4}}, 1, 2^{\frac{1}{4}}, 2^{\frac{1}{2}}, 2^{\frac{3}{4}}, 2\right\}$. And in the second-order masking experiments, the value of σ varies in the set $\left\{2^{-\frac{7}{4}}, 2^{-\frac{3}{2}}, 2^{-\frac{5}{4}}, 2^{-1}, 2^{-\frac{3}{4}}, 2^{-\frac{1}{2}}, 2^{-\frac{1}{4}}, 1\right\}$.

Experimental results The attack results are shown in Figs. 2a and 3a, respectively. With these experiments, one can note that the results of masked S-Boxes are basically consistent with those of unprotected S-Boxes, especially in the case of low noise. The S-Box of Elephant is the most resistant against CPA attacks, and the S-Boxes of Spook and SKINNY-64 are the weakest. In addition, with the noise increase, the order of S-Boxes resistance against CPA attacks fluctuates slightly in the experimental results. We argue that this is due to the increase of noise, which makes the evaluation results unstable.

Practical experiments

Experimental setup For the first- and second-order masking cases, the masking scheme proposed in Benadjila et al. (2020) and Valiveti and Vivek (2020) are adopted and implemented as our attack targets, respectively.

Experimental results The attack results are shown in Figs. 2b and 3b. It can be observed that for most S-Box examples, we obtain similar results. Namely, those S-Boxes with lower VTO_0 (higher CCV) values still have higher CPA resistance in real environments.

Profiled side-channel attacks

In this section, we further investigate the resistance of different S-Boxes against profiled side-channel attacks and check whether the three metrics are applicable to profiled attacks scenario.

Profiled side-channel attacks consist of two phases: the offline profiling phase and the online attack phase. The attacker is assumed to have an open copy of the target device to learn the leakage distribution and to perform attacks with the learned models. In profiling phase, the attacker has a device with knowledge about the secret key implemented and acquires a set of N side-channel traces $\mathcal{L}_{\text{profiling}} = \left\{\widetilde{\boldsymbol{l}}_j \mid j = 1, 2, \ldots, N\right\}$. Each trace $\widetilde{\boldsymbol{l}}_j$ is corresponding to sensitive variable $y_j = f(p_j, k)$ in one encryption (or decryption) with known key $k \in \mathcal{K}$ and plaintext (or ciphertext) p_j. Once the acquisition is done, the attacker builds suitable models and computes the estimation of probability:

$$\Pr[\boldsymbol{L} \mid Y = y], \tag{5}$$

from a profiling set $\left\{(\widetilde{\boldsymbol{l}}_j, y_j)\right\}_{j=1}^{N}$. Then in attack phase, the attacker attempts to recover the unknown key in the target device with the help of profiled leakage details.

Specifically, we launch template attacks and deep learning based profiled attacks by simulated and practical experiments.

Template attacks on the nine 4 × 4 S-Boxes

Among profiled attacks, template attack (TA) (Chari et al. 2002) and its modified version efficient template attack (ETA) (Choudary and Kuhn 2013) are the most popular

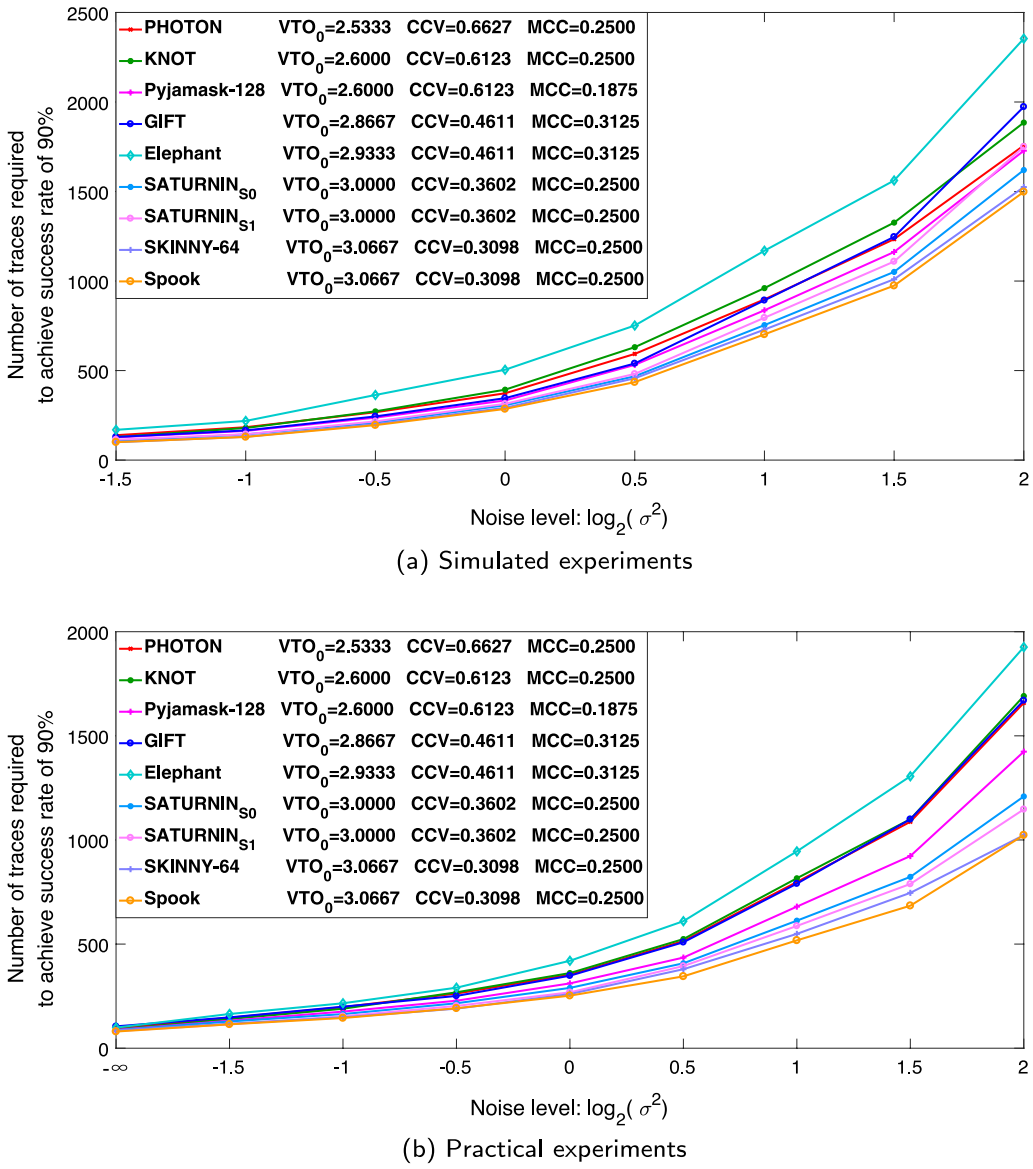
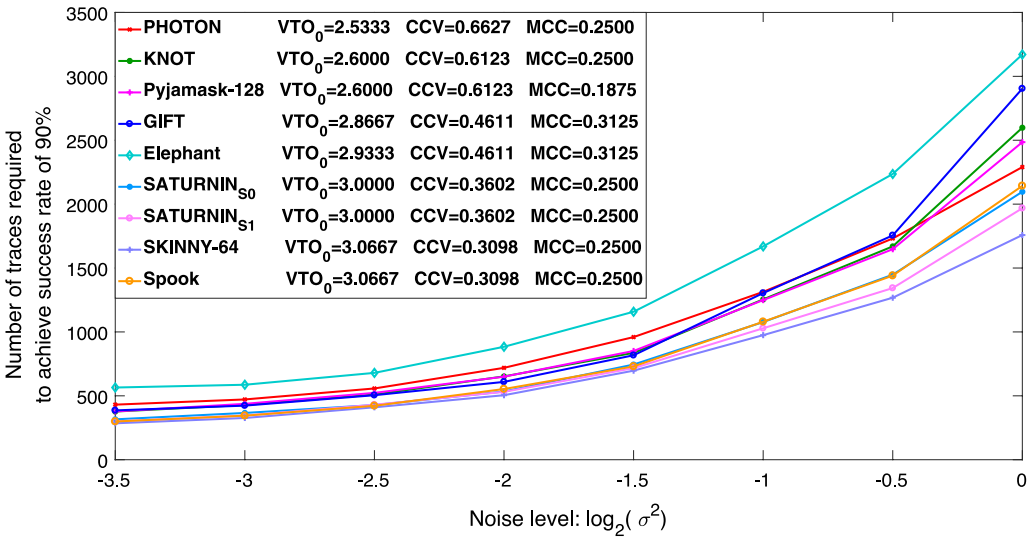


Fig. 2 CPA attacks on first-order masked 4 × 4 S-Boxes

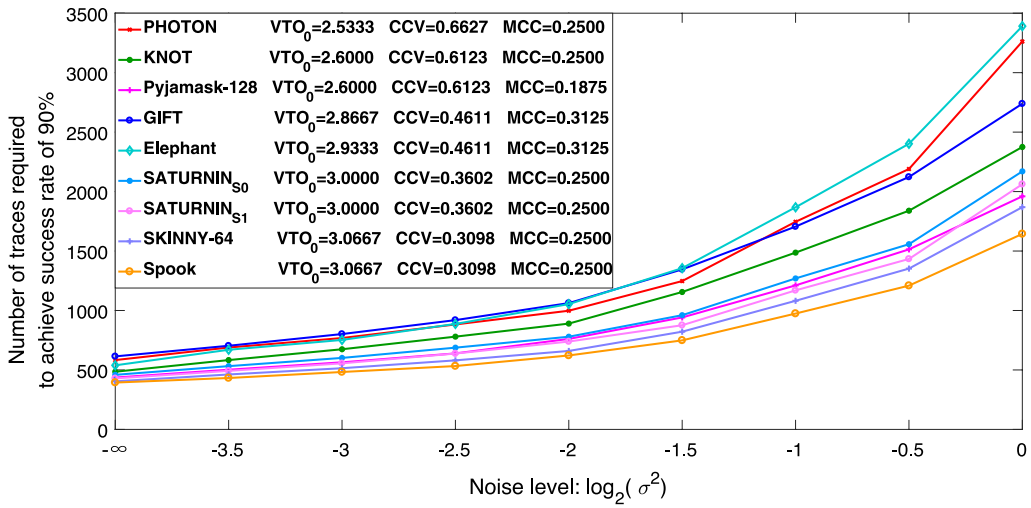
and widely used approaches. In TA, the attacker assumes that $\boldsymbol{L} \mid Y$ has a multivariate Gaussian distribution, and estimates the mean vector $\boldsymbol{\mu}_y$ and $\boldsymbol{\Sigma}_y$ for each $y \in \mathcal{Y}$ (i.e. the so-called templates). In this way, Eq. (5) is approximated by the Gaussian probability distribution function with parameters $\boldsymbol{\mu}_y$ and $\boldsymbol{\Sigma}_y$. And in ETA, the attacker replaces the covariance matrixes with one pooled covariance matrix to cope with some statistical difficulties (Choudary and Kuhn 2013). In this paper, ETA is adopted to evaluate the resistance of the S-Boxes. In the attack phase, the attacker acquires a small new set of traces $\mathcal{L}_{\text{attack}} = \{\boldsymbol{l}_j \mid j = 1, 2, \ldots, Q\}$ with a fixed unknown key k^*. With the knowledge of the established models, the estimated posterior probabilities can be calculated *via* the Bayes' Theorem. Then the attacker can select the key that maximizes the probability following the Maximum Likelihood strategy:

$$k^* = \underset{k \in \mathcal{K}}{\operatorname{argmax}} \prod_{j=1}^{Q} \frac{\Pr[\boldsymbol{L} = \boldsymbol{l}_j \mid Y = f(p_j, k)] \cdot \Pr[Y = f(p_j, k)]}{\Pr[\boldsymbol{L} = \boldsymbol{l}_j]}. \tag{6}$$

Equation (6) stands only when acquisitions are independent, which is a practical condition in reality. Notice that the attacker can launch a high-order template attack



(a) Simulated experiments



(b) Practical experiments

Fig. 3 CPA attacks on second-order masked 4 × 4 S-Boxes

if the leakages exist in high-order moments of sample points, such as defeating mask countermeasures.

Similar to the previous section, we also study the resistance of S-Boxes in unprotected, first- and second-order masking cases, respectively.

Experiments of the Unprotected S-Boxes

Experimental setup We perform both simulated and practical attacks to compare different S-Boxes. As for simulated experiments, the leakages are simulated in the same way as in the non-profiled scenario. In detail, we generate 3 points of interest (PoIs) corresponding to the output of S-Boxes. As for practical experiments, the experimental setup is exactly the same as that in the previous section, and we pre-select 3 PoIs with the highest Pearson correlation coefficient. We profile 16 efficient templates using 10,000 traces for each S-Box. And attacks are performed at almost no leakage noise, low leakage noise and high leakage noise levels ($\sigma = 0.1$, $\sigma = 1$ and $\sigma = 2$), respectively. For each S-Box, we run ETA attacks 100 times with randomly selected sub-samples of attack set for evaluation and record the minimum number of traces N required to achieve an attack success rate of 90%.

Experimental results The experimental results are shown in Fig. 14 of the Appendix. We can observe that the resistance of different unprotected S-Boxes against ETA attacks is very close, even under high noise condition. We believe the main reason is that the efficient templates have a good characterization of the leakages in both simulated and practical experiments. Therefore, we further investigate the resistance of different S-Boxes in first- and second-order masking cases.

Experiments of the masked S-Boxes

In the profiling phase, we first profile 16 efficient templates using 10,000 traces for each share. Next, in the attack phase, we match the leakages to the profiled templates, which are denoted as $\boldsymbol{M^i}$ and $i \in \{0, 1, \ldots, d\}$. Then we get the probability $P\left(Y_j^i = y_j^i \mid \boldsymbol{l}_j^i, \boldsymbol{M^i}\right)$ utilizing the efficient templates for each trace. Where y_j^i denotes the i-th share of the output of the S-Box corresponding to the j-th trace, and $\boldsymbol{l}_j^i$ denotes the leakage for the i-th share of the j-th trace. The probability of y_j can be expressed as:

$$P(y_j \mid \boldsymbol{l}_j, \boldsymbol{M}) = \sum_{\mathcal{S}} \prod_{j=0}^{d} P\left(y_j^i \mid \boldsymbol{l}_j^i, \boldsymbol{M^i}\right),$$

where $\mathcal{S}$ is the set $\left\{\left(y_j^0, \ldots, y_j^d\right) \mid y_j = y_j^0 \oplus \cdots \oplus y_j^d\right\}$, and $\boldsymbol{l}_j$ denotes the leakages of all shares of the j-th trace. With the information of the inverse mapping and the plaintext, $P(y_j)$ can be mapped to $P_j(k)$. Add up the $P_j(k)$ of all the attack traces, and the key hypothesis corresponding to the maximum value of $P(k)$ is the revealed key.

Experimental setup As for simulated experiments, we generate 3 PoIs corresponding to each share of the output of S-Boxes. As for practical experiments, we also preselect 3 PoIs for each share to construct templates and perform attacks. The remaining experimental settings are the same as those in the previous experiments.

Experimental results The attack results of first- and second-order masking cases with different noise levels are shown in Figs. 4 and 5, respectively. As for the second-order masking case, the increase of noise will seriously affect the stability of the attack results and the accuracy of the evaluation, so we only show the experimental results when $\sigma = 0.1$ and $\sigma = 1$. It can be observed that in both first- and second-order masking implementations, when the noise level is very low, the resistance of different S-Boxes against ETA attacks is still very close to each other. So we think that in very low-noise scenarios, it doesn't seem necessary to consider how to select optimal 4×4 S-Boxes against ETA attacks.

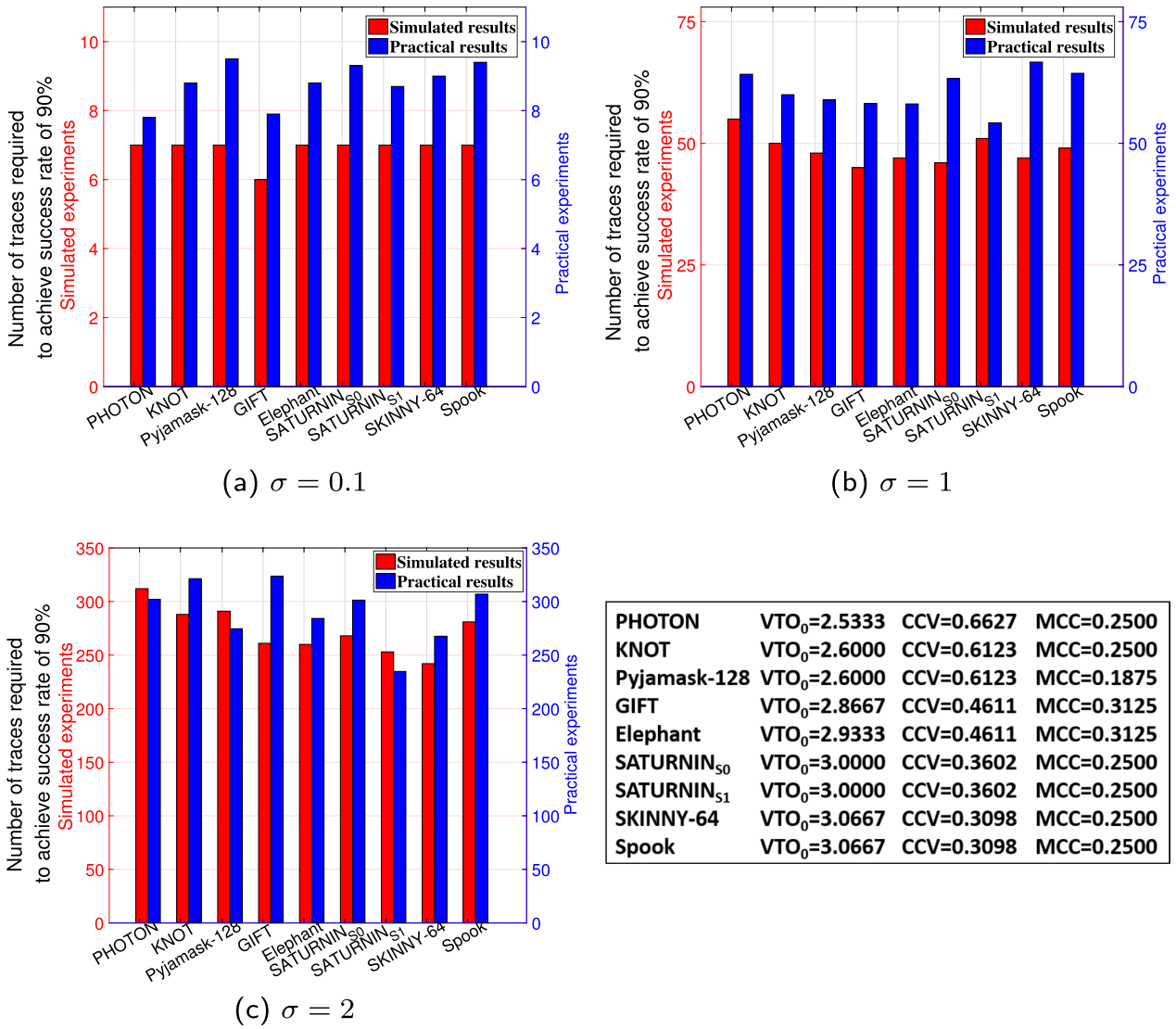
With the noise increase, the difference between different S-Boxes becomes slightly more significant. However, the practical results are not consistent with the simulation results. We infer the main reasons for the inconsistent results are the leakages in the real environment do not fully satisfy the HW leakage model and the noise does not fulfill the Gaussian noise assumption. And with the noise increase, the accuracy of the constructed templates is seriously affected. In addition, neither the simulated results nor the practical results are consistent with the results of all the three metrics. We argue that this is because the characterization of the noise, rather than the intrinsic properties of S-Boxes, is the dominant factor affecting the effectiveness of the attacks. Therefore, these metrics may not be suitable for evaluating the resistance of S-Boxes against template attacks.

In addition, we find that the difference between S-Boxes against ETA is far less than that of S-Boxes against CPA attacks. And the experimental results of ETA are not consistent with those of CPA attacks. For example, the S-Box of Elephant is the most resistant to CPA attacks, but obviously not the most resistant to ETA attacks. And none of the 4-bit S-Boxes shows significantly more resistant than the others. We also perform attacks that target the HW of the outputs of the S-Boxes (profiling 5 efficient templates), again with no clear pattern that could be observed. The possible reason is that the intrinsic properties of the S-Boxes we analyzed are relatively close to each other. Whatever, when selecting the optimal S-Boxes, it is necessary to comprehensively consider the resistance of S-Boxes against a different type of attacks. It is not sufficient to consider only transparency orders or confusion coefficients.

Deep learning based profiled attacks

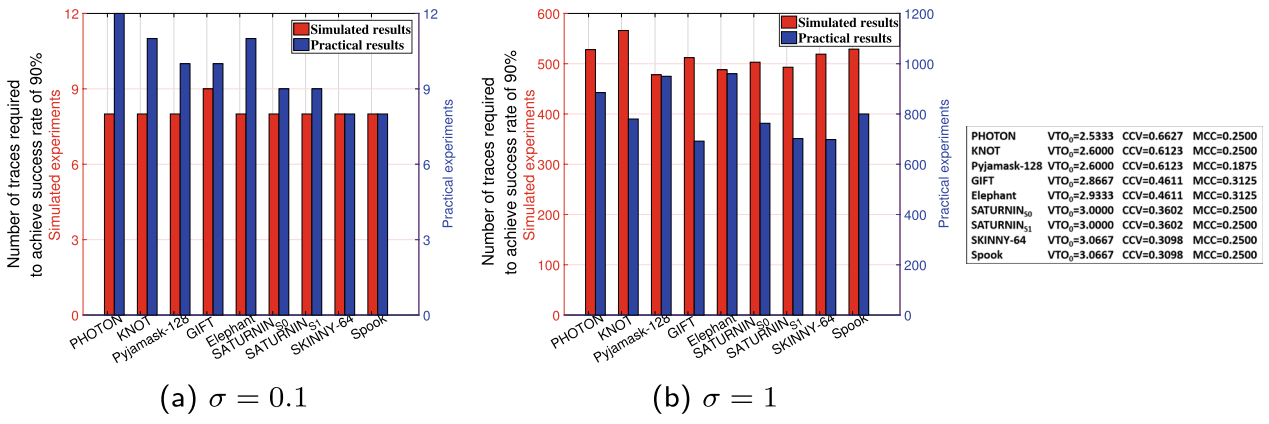
Recently, deep learning techniques gained substantial interest in the community of side-channel analysis. Previous researches have evidenced deep learning based attacks give a very efficient alternative to the state-of-the-art profiled attacks, and even outperform the traditional profiled attacks (Maghrebi et al. 2016; Cagli et al. 2017). We explore the resistance of the nine 4×4 S-Boxes against such attacks, and whether the three metrics are effective when measuring the resistance against deep learning based attacks. According to the work in Wouters et al. (2020), when the traces are synchronized, the Multi Layer Perceptron (MLP) models are as effective as Convolutional Neural Network (CNN) models. Since we only consider the case of the traces are aligned in this work, the attacks based on the MLP networks are performed.

In this subsection, all experiments are conducted on an Intel(R) Xeon(R) CPU E5-2667 v4 @3.20 GHz 32 core machine with two NVIDIA TITAN Xp GPUs. We use the



PHOTON	VTO$_0$=2.5333	CCV=0.6627	MCC=0.2500
KNOT	VTO$_0$=2.6000	CCV=0.6123	MCC=0.2500
Pyjamask-128	VTO$_0$=2.6000	CCV=0.6123	MCC=0.1875
GIFT	VTO$_0$=2.8667	CCV=0.4611	MCC=0.3125
Elephant	VTO$_0$=2.9333	CCV=0.4611	MCC=0.3125
SATURNIN$_{S0}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SATURNIN$_{S1}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SKINNY-64	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500
Spook	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500

Fig. 4 Simulated and practical ETA attacks on first-order masked 4 × 4 S-Boxes



PHOTON	VTO$_0$=2.5333	CCV=0.6627	MCC=0.2500
KNOT	VTO$_0$=2.6000	CCV=0.6123	MCC=0.2500
Pyjamask-128	VTO$_0$=2.6000	CCV=0.6123	MCC=0.1875
GIFT	VTO$_0$=2.8667	CCV=0.4611	MCC=0.3125
Elephant	VTO$_0$=2.9333	CCV=0.4611	MCC=0.3125
SATURNIN$_{S0}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SATURNIN$_{S1}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SKINNY-64	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500
Spook	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500

Fig. 5 Simulated and practical ETA attacks on second-order masked 4 × 4 S-Boxes

Keras library (version 2.2.2) with the TensorFlow library (version 1.10.0) as the backend for MLP.

MLP architecture We refer to the recent work (Wouters et al. 2020) and then design our MLP models. For the unprotected and first-order masking cases, the MLP is composed of one hidden layer with 10 neurons. And for the second-order masking case, the MLP is composed of two hidden layers with 10 neurons. Each layer is activated by the ReLU function and He Uniform initialization is used to improve the weight initialization. The output layer contains 16 neurons activated by the softmax function. Cross-entropy is used as the loss function. As a remark, the network architectures used in this subsection are surely not optimal, as our goal is not to select the optimal parameters.

For the training of MLP networks, the mini-batch size is 128 and the maximum iterative epoch is 100. And the network kernel weights are recorded for the best validation loss. Once the training is done, we reconstruct the neuron network with the best recorded weights. The learning rate is initially 0.005, and a technique called One Cycle Policy (Smith 2017) is used to choose the right learning rate.

Experiments of the Unprotected S-Boxes

Experimental setup As for simulated experiments, we generate 10 sample points for each trace, of which the first three points are PoIs corresponding to the output of S-Boxes and the rest are randomly generated in [0, 4]. As for practical experiments, 10 samples that contain information on the output of S-Boxes are captured for each trace. There are 10,000 traces for profiling and 5,000 traces for the attack. In the profiling traces, 90% are used for training and 10% are used for validation. We run each attack 100 times with randomly selected sub-samples of attack sets and record the minimum number of traces required to achieve an attack success rate of 90%. Since the training of the neural network might be unstable, we repeat the experiments 10 times and take the average results.

Experimental results The experimental results are shown in Fig. 15 of the Appendix. Similar to the results of ETA attacks, the resistance of different unprotected S-Boxes against deep learning based attacks is still very close, even under the high noise condition. Next, we further investigate the resistance of different S-Boxes in first- and second-order masking cases.

Experiments of the masked S-Boxes

Experimental setup Both the simulated and practical traces consist of 10 sample points. As for simulated experiments, we generate 3 PoIs corresponding to each share of the output of S-Boxes, and the rest are randomly generated in [0, 4]. As for practical experiments, 10 samples that contain information on each share of the output of S-Boxes are captured. For the first-order masking case, there are 10,000 traces for profiling and 10,000 traces for the attack. And for the 2nd-order masking case, there are 30,000 traces for profiling and 20,000 traces for the attack.

Experimental results The results of first- and second-order masking cases are shown in Figs. 6 and 7, respectively. Similar to the results of ETA attacks, when the noise level is very low, the resistance of different S-Boxes against deep learning based attacks is still very close to each other in both first- and second-order masking cases. As the noise increases, the difference between different S-Boxes becomes more obvious. However, we still cannot find patterns in the experimental results. On the one hand, the practical results are not consistent with the simulation results. In addition to the reasons mentioned above, the instability of the network training may also contribute to this phenomenon. On the other hand, neither the simulated results nor the practical results are consistent with the results of all the three metrics. Namely, all the three metrics are not suitable for evaluating the resistance of S-Boxes against deep learning based attacks. Therefore, how to quantify the resistance of S-Boxes against deep learning based attacks still has a long way to go.

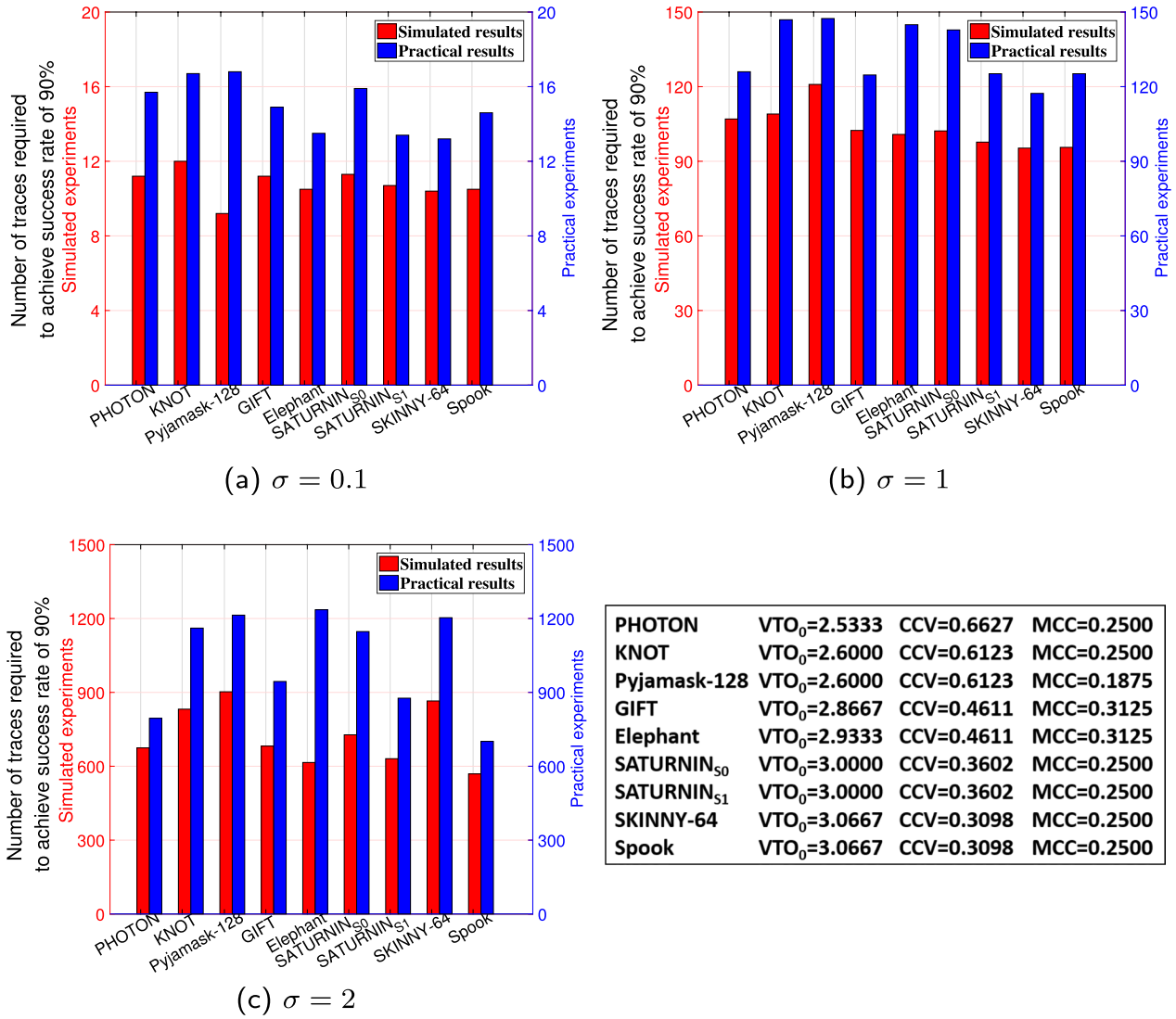
4×4 S-Boxes versus 8×8 S-Boxes

In this section, taking several 4×4 S-Boxes and 8×8 S-Boxes as examples, we verify whether VTO, CCV and MCC can be applied to compare the resistance of S-Boxes with different sizes through simulated and practical experiments.

Non-profiled side-channel attacks

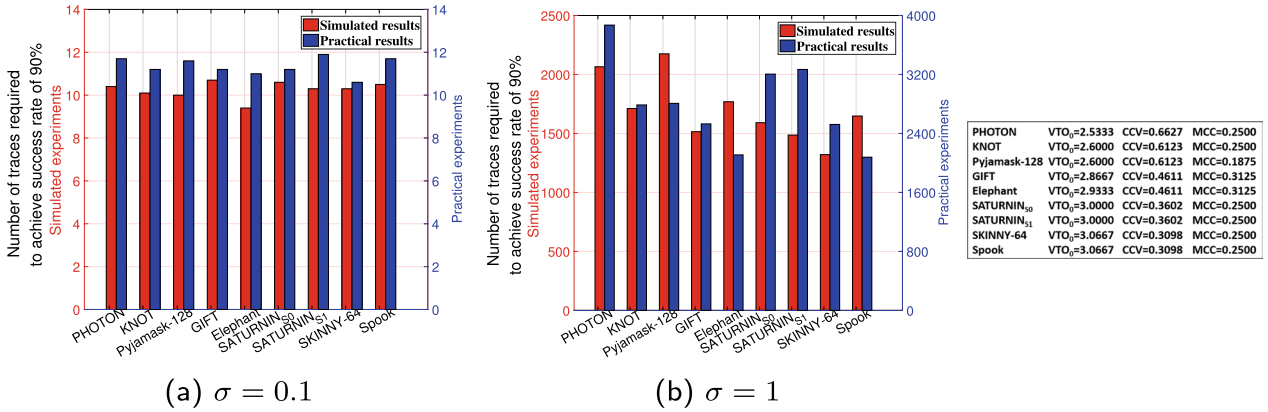
From the perspective of theoretical analysis, among nine 4×4 S-Boxes, the S-Box of PHOTON is the hardest to attack, and the S-Box of Spook is one of the easiest to attack. In addition, according to the experimental results, the S-Box of Elephant is the most resistant against CPA attacks, and the S-Box of Spook is one of the easiest to attack. Considering the above factors, we select the S-Boxes of PHOTON, Elephant, and Spook as the representatives of the 4×4 S-Boxes to compare with the 8×8 S-Boxes of SKINNY-128 and AES.

Experimental setup We study the resistance of S-Boxes in unprotected, first- and second-order masking cases, respectively. And the simulated and practical experiments are performed with different noise levels. Due to the simulated traces and practical traces are standardized (zero mean and unit variance) before Gaussian noise



PHOTON	VTO$_0$=2.5333	CCV=0.6627	MCC=0.2500
KNOT	VTO$_0$=2.6000	CCV=0.6123	MCC=0.2500
Pyjamask-128	VTO$_0$=2.6000	CCV=0.6123	MCC=0.1875
GIFT	VTO$_0$=2.8667	CCV=0.4611	MCC=0.3125
Elephant	VTO$_0$=2.9333	CCV=0.4611	MCC=0.3125
SATURNIN$_{S0}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SATURNIN$_{S1}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SKINNY-64	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500
Spook	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500

Fig. 6 Simulated and practical deep learning based profiled attacks on first-order masked 4 × 4 S-Boxes



PHOTON	VTO$_0$=2.5333	CCV=0.6627	MCC=0.2500
KNOT	VTO$_0$=2.6000	CCV=0.6123	MCC=0.2500
Pyjamask-128	VTO$_0$=2.6000	CCV=0.6123	MCC=0.1875
GIFT	VTO$_0$=2.8667	CCV=0.4611	MCC=0.3125
Elephant	VTO$_0$=2.9333	CCV=0.4611	MCC=0.3125
SATURNIN$_{S0}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SATURNIN$_{S1}$	VTO$_0$=3.0000	CCV=0.3602	MCC=0.2500
SKINNY-64	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500
Spook	VTO$_0$=3.0667	CCV=0.3098	MCC=0.2500

Fig. 7 Simulated and practical deep learning based profiled attacks on second-order masked 4 × 4 S-Boxes

added, the 4 × 4 S-Boxes and 8 × 8 S-Boxes are compared at almost the same SNR.

Experimental results The results of simulated and practical experiments are shown in Figs. 8 and 9, respectively. In the unprotected case, we can observe that the S-Boxes of SKINNY-128 and AES perform worse than that of Elephant, similar to PHOTON, and better than Spook. Therefore, the 4 × 4 S-Boxes that are selected carefully could be even more resistant against CPA attacks than certain 8 × 8 S-Boxes. However, according to the values of theoretical metrics, the two 8 × 8 S-Boxes lead to higher values of VTO_0 and MCC than the 4 × 4 S-Boxes, which implies 8 × 8 S-Boxes are more vulnerable to attacks. And the resistance of the S-Box of SKINNY-128 should be worse than that of PHOTON and Elephant, and slightly better than that of Spook in terms of VCC. As for the S-Box of AES, it should be the easiest to attack among all the S-Boxes. The inconsistency between theoretical analysis and practical results indicates that none of the three metrics can be used to quantify and compare S-Boxes with different sizes.

As for first- and second-order masking cases, the two 8 × 8 S-Boxes perform much better than the 4 × 4 S-Boxes. The main reason is that the 8-bit masks provide much better randomization than the 4-bit masks. Of course, the larger size of S-Boxes also leads to higher implementation costs. This is a trade-off between the security and costs, which is outside the scope of this work.

In addition, for the two 8 × 8 S-Boxes we evaluated, the S-Box of SKINNY-128 always performs better than that of AES. However, the results in Heuser et al. (2016) show that the 4 × 4 S-Boxes they studied have a different side-channel resiliency, while the difference in the 8 × 8 S-Boxes is only theoretically present. We argue that a good selection of 8 × 8 S-Boxes could also result in an improvement in inherent resilience.

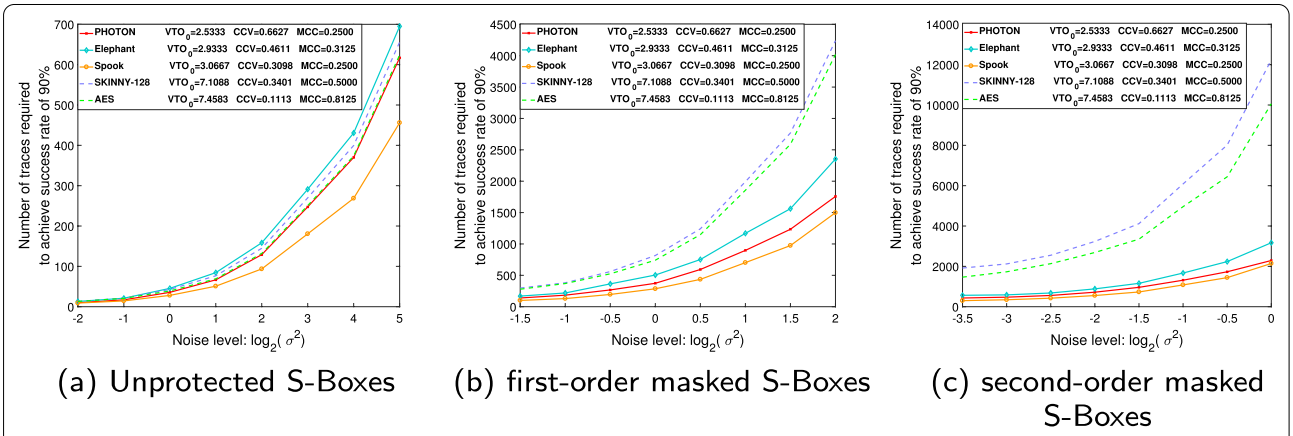


Fig. 8 Simulated CPA attacks on 4 × 4 S-Boxes and 8 × 8 S-Boxes

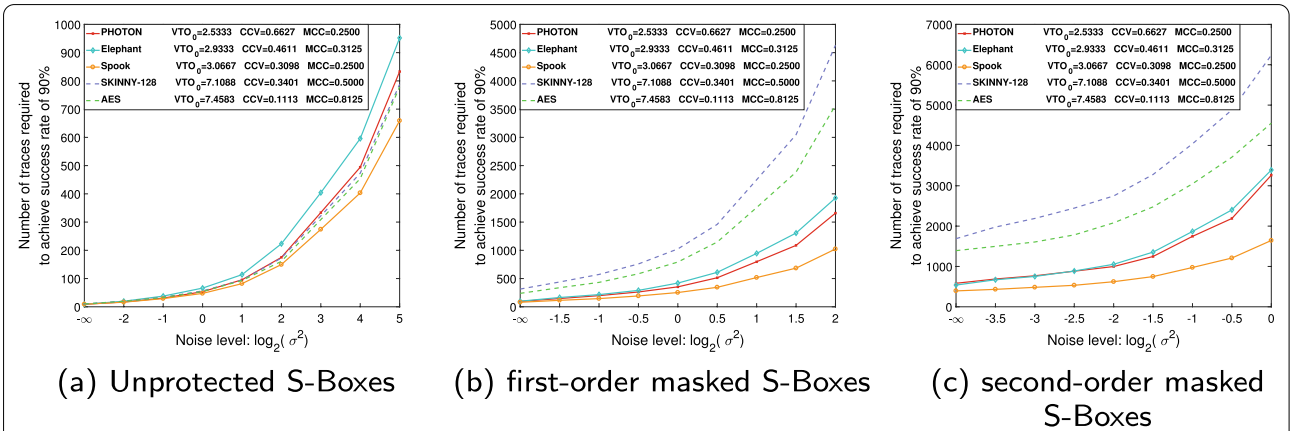


Fig. 9 Practical CPA attacks on 4 × 4 S-Boxes and 8 × 8 S-Boxes

Profiled side-channel attacks

In this section, we compare the resistance of 4-bit and 8-bit S-Boxes against profiled side-channel attacks. The S-Boxes used are the same as above.

Template attacks

Experimental setup We study the resistance of S-Boxes in unprotected and first-order masking cases. And the simulated and practical experiments are performed with different noise levels. We profile 16 efficient templates using 10,000 traces for each 4×4 S-Box, and profile 256 efficient templates using 160,000 traces for each 8×8 S-Box. Therefore, the number of profiling traces for each class of 4×4 S-Boxes and 8×8 S-Boxes is roughly the same.

Experimental results The results of the unprotected and first-order cases are shown in Figs. 10 and 11, respectively. In the unprotected case, we can observe that the resistance of 8-bit S-Boxes and 4-bit S-Boxes are quite close. The main reason is that the efficient templates have a good characterization of the leakages. As for the first-order case, it is obvious that the two 8-bit S-Boxes are more resistant against ETA attacks than the 4-bit S-Boxes. It seems natural since 8-bit S-Boxes have a significantly larger number of classes than 4-bit S-Boxes. In addition, in practical experiments, the difference between the 4-bit and 8-bit S-Boxes is larger than that in the simulated experiments. We infer the main reasons are the leakages in the real environment do not fully satisfy the HW leakage model and the noise does not fulfill the Gaussian noise assumption. Because the traces of the 8-bit S-Box is divided into 256 classes, it requires higher precision of the constructed templates, and then the accuracy decreases faster.

Deep Learning Based Profiled Attacks

Experimental setup We study the resistance of 4×4 S-Boxes and 8×8 S-Boxes against deep learning based profiled attacks. The simulated and practical experiments are performed with different noise levels. We profile 16 efficient templates using 10,000 traces for each 4×4 S-Box, and profile 256 efficient

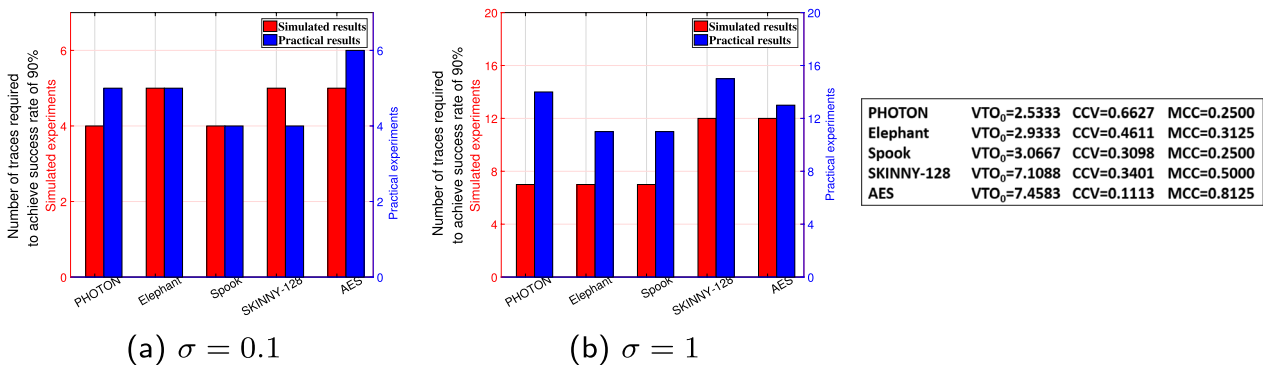


Fig. 10 Simulated and practical ETA attacks on unprotected 4×4 S-Boxes and 8×8 S-Boxes

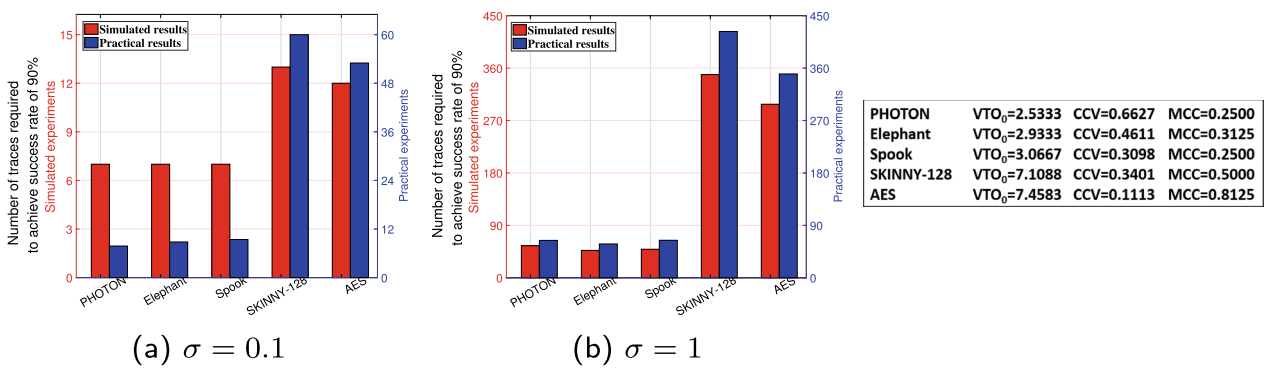


Fig. 11 Simulated and practical ETA attacks on first-order masked 4×4 S-Boxes and 8×8 S-Boxes

templates using 160,000 traces for each 8×8 S-Box. The network architectures and other experimental settings are the same as those in the previous section.

Experimental results The results of the unprotected and first-order cases are shown in Figs. 12 and 13, respectively. It is obvious that, in both unprotected and first-order cases, the two 8-bit S-Boxes are more resistant against deep learning based profiled attacks than the 4-bit S-Boxes. It implies that, when the leakages cannot be characterized very accurately, S-Boxes with larger sizes are more resistant than S-Boxes with smaller sizes. Interestingly, for the first-order case, practical attacks perform even better than simulated attacks. We guess the reason is the irregular noise in practical traces alleviates the overfitting during the training of networks. This phenomenon also shows that when evaluating the resistance of S-Boxes against deep learning based side-channel attacks, it is not sufficient to perform simulated experiments alone.

Conclusions and future work

In this paper, taking the S-Boxes used in NIST Lightweight Cryptography candidates as concrete examples, we give a comprehensive study of the applicability of three popular theoretical metrics for side-channel analysis, namely VTO, CCV and MCC. Firstly, we find that CCV is almost linearly correlated with VTO, while MCC is inconsistent with the other two metrics. Next, to verify which metric is more effective in which scenarios, we perform simulated and practical experiments on nine 4-bit S-Boxes in the non-profiled and profiled scenarios, respectively. For the non-profiled attacks, when the difference of VTO (resp. CCV) values of the two S-Boxes is relatively large, the S-Box with a lower VTO (resp. higher CCV) value is generally more resistant to CPA attacks. However, when VTO and CCV values of S-Boxes become relatively close to each other, these two metrics turn less accurate. Interestingly, MCC fails to work in quantifying the resistance of S-Boxes against CPA attacks. As for the

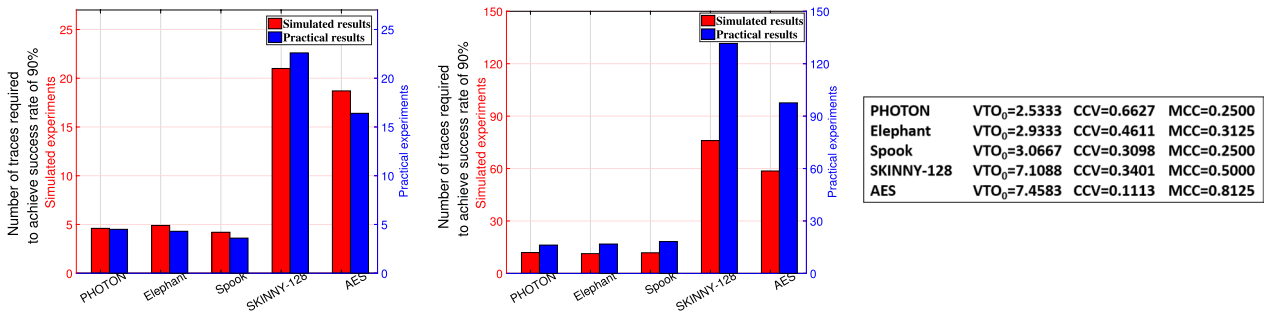


Fig. 12 Simulated and practical deep learning based profiled attacks on unprotected 4×4 S-Boxes and 8×8 S-Boxes

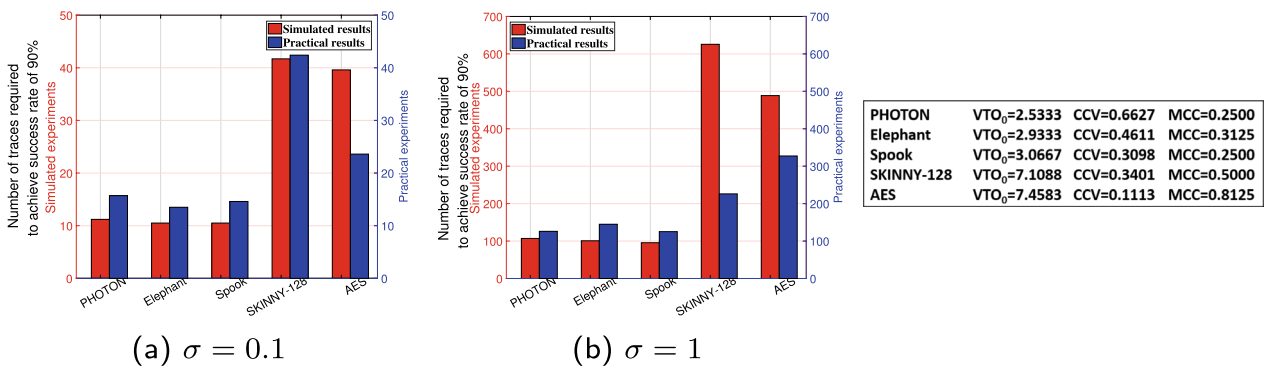


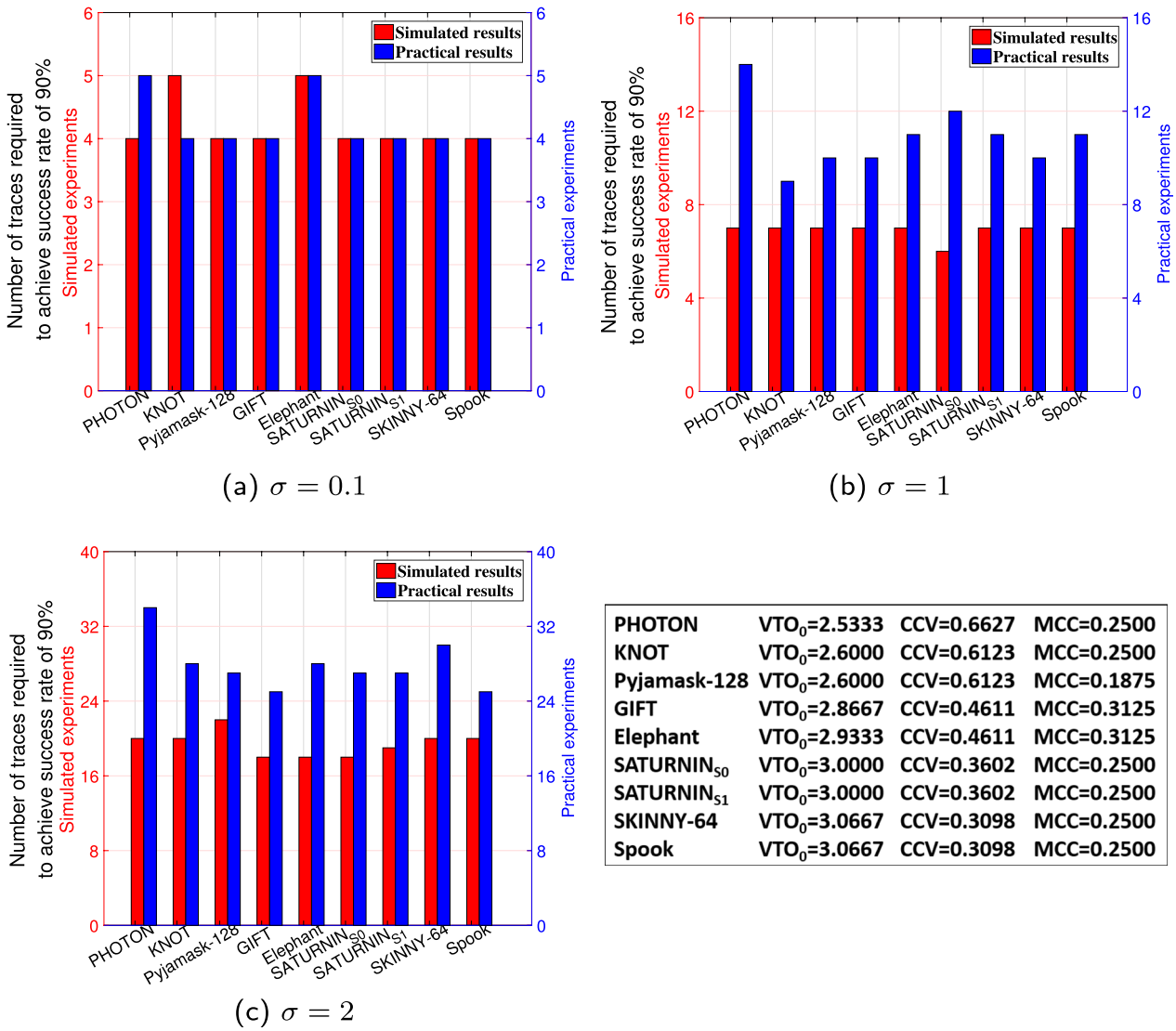
Fig. 13 Simulated and practical deep learning based profiled attacks on first-order masked 4×4 S-Boxes and 8×8 S-Boxes

profiled scenario, we perform efficient template attacks and deep learning based profiled attacks. However, none of the three metrics is suitable for measuring the resistance of S-Boxes against profiled SCAs. Finally, we try to verify whether these metrics can be applied to compare the resistance of S-Boxes with different sizes. Unfortunately, all the three metrics fail to work when measuring and comparing S-Boxes with different sizes.

Since VTO and CCV lack the accuracy to evaluate the resistance of S-Boxes against CPA-like attacks, it is significant to further analyze the reasons for the lack of precision of the existing metrics, and then explore the theoretical metric that fits the reality better. Additionally, exploring the theoretical relationship between transparency order and confusion coefficients may be helpful to propose the new metric.

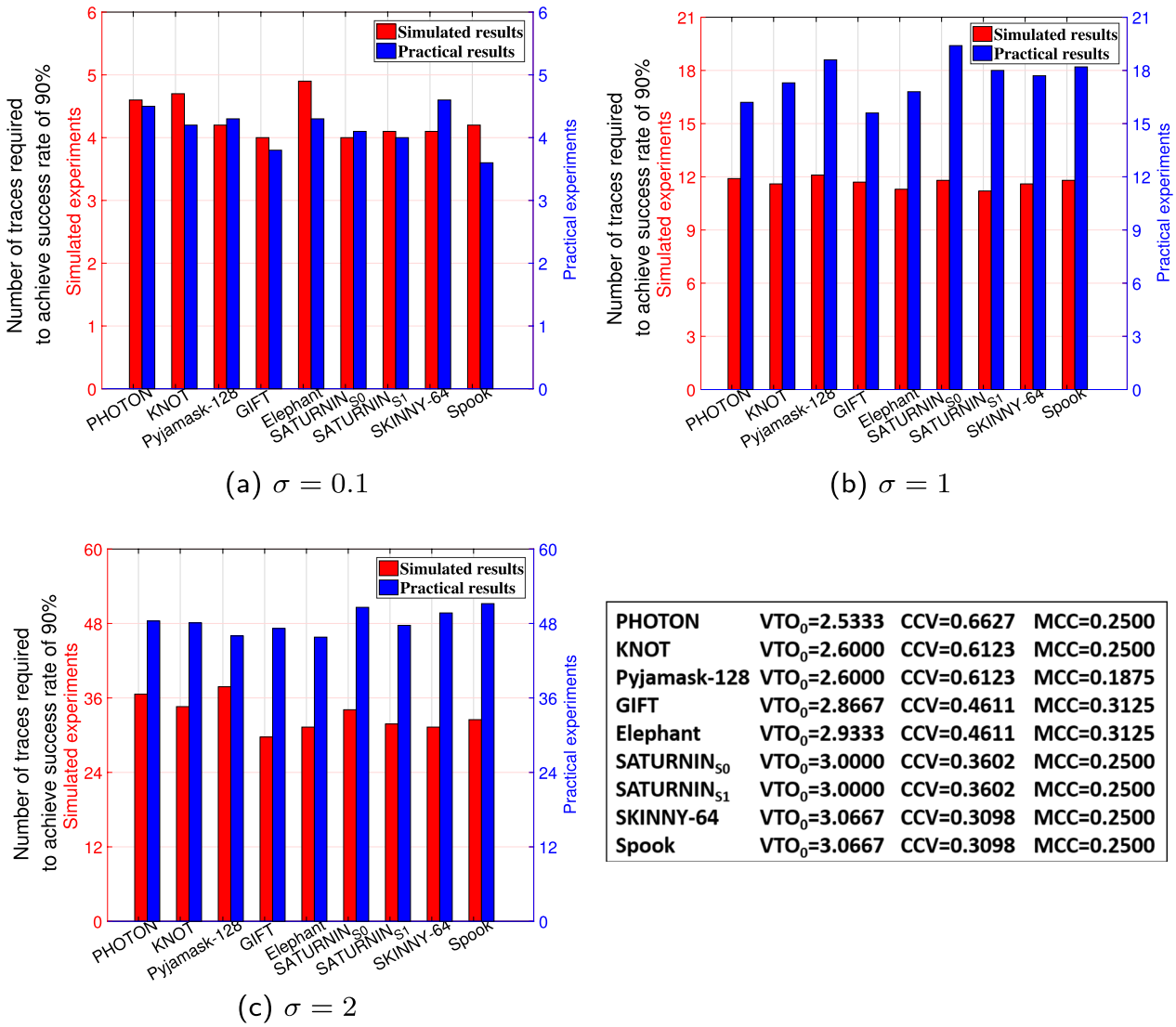
Appendix

The experimental results of nine 4×4 unprotected S-Boxes against efficient template attacks and deep learning based profiled attacks are shown in Figs. 14 and 15, respectively.



PHOTON	VTO_0=2.5333	CCV=0.6627	MCC=0.2500
KNOT	VTO_0=2.6000	CCV=0.6123	MCC=0.2500
Pyjamask-128	VTO_0=2.6000	CCV=0.6123	MCC=0.1875
GIFT	VTO_0=2.8667	CCV=0.4611	MCC=0.3125
Elephant	VTO_0=2.9333	CCV=0.4611	MCC=0.3125
SATURNIN$_{S0}$	VTO_0=3.0000	CCV=0.3602	MCC=0.2500
SATURNIN$_{S1}$	VTO_0=3.0000	CCV=0.3602	MCC=0.2500
SKINNY-64	VTO_0=3.0667	CCV=0.3098	MCC=0.2500
Spook	VTO_0=3.0667	CCV=0.3098	MCC=0.2500

Fig. 14 Simulated and practical ETA attacks on unprotected 4×4 S-Boxes



PHOTON	VTO_0=2.5333	CCV=0.6627	MCC=0.2500
KNOT	VTO_0=2.6000	CCV=0.6123	MCC=0.2500
Pyjamask-128	VTO_0=2.6000	CCV=0.6123	MCC=0.1875
GIFT	VTO_0=2.8667	CCV=0.4611	MCC=0.3125
Elephant	VTO_0=2.9333	CCV=0.4611	MCC=0.3125
SATURNIN$_{S0}$	VTO_0=3.0000	CCV=0.3602	MCC=0.2500
SATURNIN$_{S1}$	VTO_0=3.0000	CCV=0.3602	MCC=0.2500
SKINNY-64	VTO_0=3.0667	CCV=0.3098	MCC=0.2500
Spook	VTO_0=3.0667	CCV=0.3098	MCC=0.2500

Fig. 15 Simulated and practical deep learning based profiled attacks on unprotected 4 × 4 S-Boxes

Acknowledgements
Not applicable.

Authors' contributions
HL completed the main work of the paper and drafted the manuscript. GY and JM participated in the experiments of non-profiled and profiled SCAs, respectively. YZ participated in problem discussions and improvements of the manuscript. CJ participated in the experiments of SCAs against 8-bit S-boxes. All authors read and approved the final manuscript.

Funding
This work is supported in part by National Natural Science Foundation of China (Nos. 61632020, U1936209 and 62002353) and Beijing Natural Science Foundation (No.4192067).

Availability of data and materials
Not applicable.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China. [2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China. [3]School of Cyber Security, Nanjing University of Science and Technology, Nanjing 210094, China. [4]Department of Software Engineering and Software Evaluation, Beijing Institute of Computer Technology and Application, Beijing 100039, China.

Received: 2 August 2021 Accepted: 7 October 2021
Published online: 01 November 2021

References

Andreeva E, Lallemand V, Purnal A, Reyhanitabar R, Roy A, Vizár D (2019) Forkae v. In: Submission to NIST lightweight cryptography project

Banik S, Bogdanov A, Peyrin T, Sasaki Y, Sim SM, Tischhauser E Todo Y (2019) Sundae-gift. In: Submission to NIST lightweight cryptography project **1**

Banik S, Chakraborti A, Iwata T, Minematsu K, Nandi M, Peyrin T, Sasaki Y, Sim SM, Todo Y (2019) Gift-cofb. In: Submission to NIST lightweight cryptography project **1**

Bao Z, Chakraborti A, Datta N, Guo J, Nandi M, Peyrin T, Yasuda K (2019) Photon-beetle authenticated encryption and hash family. Submiss NIST Lightweight Cryptogr Proj 1:115

Beierle C, Jean J, Kölbl S, Leander G, Moradi A, Peyrin T, Sasaki Y, Sasdrich P, Sim SM (2020) SKINNY-AEAD and skinny-hash. IACR Trans Symmetric Cryptol 2020(S1):88–131. https://doi.org/10.13154/tosc.v2020.iS1.88-131

Bellizia D, Berti F, Bronchain O, Cassiers G, Duval S, Guo C, Leander G, Leurent G, Levi I, Momin C et al (2020) Spook: Sponge-based leakage-resistant authenticated encryption with a masked tweakable block cipher. IACR Trans Symmetric Cryptol 2020(S1):295–349. https://doi.org/10.13154/tosc.v2020.iS1.295-349

Benadjila R, Prouff E, Strullu R, Cagli E, Dumas C (2020) Deep learning for side-channel analysis and introduction to ASCAD database. J Cryptogr Eng 10(2):163–188. https://doi.org/10.1007/s13389-019-00220-8

Brier E, Clavier C, Olivier F (2004) Correlation power analysis with a leakage model. In: Cryptographic hardware and embedded systems—CHES 2004: 6th international workshop Cambridge, vol 3156. MA, USA, August 11–13, 2004. Springer, Berlin, pp 16–29

Cagli E, Dumas C, Prouff E (2017) Convolutional neural networks with data augmentation against jitter-based countermeasures—profiling attacks without pre-processing. In: Fischer W, Homma N (eds) Cryptographic hardware and embedded systems—CHES 2017—19th international conference, Taipei, Taiwan, September 25–28, 2017, vol 10529. Lecture Notes in Computer Science. Springer, Berlin, pp 45–68

Canteaut A, Duval S, Leurent G, Naya-Plasencia M, Perrin L, Pornin T, Schrottenloher A (2019) Saturnin: a suite of lightweight symmetric algorithms for post-quantum security

Carlet C (2005) On highly nonlinear s-Boxes and their inability to thwart DPA attacks. In: Progress in cryptology—INDOCRYPT 2005. 6th international conference on cryptology in India, Bangalore, India, December 10–12, 2005. Springer, Berlin, pp 49–62

Carlet C, de Chérisey É, Guilley S, Kavut S, Tang D (2021) Intrinsic resiliency of S-Boxes against side-channel attacks-best and worst scenarios. IEEE Trans Inf Forensics Secur 16:203–218. https://doi.org/10.1109/TIFS.2020.3006399

Chakraborti A, Datta N, Jha A, Mancillas-López C, Nandi M, Sasaki Y (2020) Estate: a lightweight and low energy authenticated encryption mode. IACR Trans Symmetric Cryptol 2020(S1):350–389. https://doi.org/10.13154/tosc.v2020.iS1.350-389

Chakraborti A, Datta N, Jha A, Lopez CM, Nandi M, Sasaki Y (2019) Lotus-aead and locus-aead. In: Submission to NIST lightweight cryptography project

Chakraborti A, Datta N, Jha A, Nandi M (2019) Hyena. In: Submission to NIST lightweight cryptography project

Chakraborty K, Sarkar S, Maitra S, Mazumdar B, Mukhopadhyay D, Prouff E (2017) Redefining the transparency order. Des Codes Crypt 82(1–2):95–115. https://doi.org/10.1007/s10623-016-0250-3

Chakraborty B, Nandi M (2019) mixFeed https://csrc.nist.gov/projects/lightweight-cryptography/round-2-candidates

Chakraborty B, Nandi M (2019) Orange. In: Submission to NIST lightweight cryptography project

Chari S, Rao JR, Rohatgi P (2002) Template attacks. In: Jr., B.S.K., Koç, Ç.K., Paar, C. (eds.) Cryptographic hardware and embedded systems—CHES 2002, 4th international workshop, Redwood Shores, CA, USA, August 13–15, 2002. Lecture Notes in Computer Science, vol. 2523. Springer, Berlin, pp. 13–28

Choudary O, Kuhn MG (2013) Efficient template attacks. In: Francillon, A., Rohatgi, P. (eds.) Smart card research and advanced applications—12th international conference, CARDIS 2013, Berlin, Germany, November 27–29, 2013. Lecture Notes in Computer Science, vol. 8419. Springer, Berlin, pp. 253–270. https://doi.org/10.1007/978-3-319-08302-5_17

de la Cruz Jiménez RA (2018) On some methods for constructing almost optimal s-boxes and their resilience against side-channel attacks. IACR Cryptol ePrint Arch 2018:618

Dobraunig C, Mennink B (2019) Elephant v1. In: Submission to NIST lightweight cryptography project

Doget J, Prouff E, Rivain M, Standaert F-X (2011) Univariate side channel attacks and leakage modeling. J Cryptogr Eng 1(2):123. https://doi.org/10.1007/s13389-011-0010-2

Duc A, Dziembowski S, Faust S (2019) Unifying leakage models: from probing attacks to noisy leakage. J Cryptol 32(1):151–177. https://doi.org/10.1007/s00145-018-9284-1

Ege B, Papagiannopoulos K, Batina L, Picek S (2015) Improving DPA resistance of S-Boxes: How far can we go? In: 2015 IEEE international symposium on circuits and systems. ISCAS 2015, Lisbon, Portugal, May 24–27, 2015. IEEE Press, Piscataway, NJ, pp 2013–2016

Fei Y, Luo Q, Ding AA (2012) A statistical model for DPA with novel algorithmic confusion analysis. In: Cryptographic hardware and embedded systems—CHES 2012—14th international workshop, vol 7428. Leuven, Belgium, September 9–12, 2012. Springer, Berlin, pp 233–250

FIPS PUB 197: Advanced encryption standard. National Institute of Standards and Technology, Gaithersburg, Maryland, USA (2001)

Freyre-Echevarría A, Martínez-Díaz I, Legón-Pérez CM, Gómez GS, Rojas O (2020) Evolving nonlinear S-Boxes with improved theoretical resilience to power attacks. IEEE Access 8:202728–202737. https://doi.org/10.1109/ACCESS.2020.3035163

Goudarzi D, Jean J, Kölbl S, Peyrin T, Rivain M, Sasaki Y, Sim SM (2019) Pyjamask v1. 0. In: Submission to NIST lightweight cryptography project

Gueron S, Jha A, Nandi M (2019) Comet: counter mode encryption with authentication tag. In: Submission to NIST lightweight cryptography project

Guilley S, Heuser A, Rioul O (2015) A key to success—success exponents for side-channel distinguishers. In: Biryukov A, Goyal V (eds) Progress in cryptology—INDOCRYPT 2015—16th international conference on cryptology in India, Bangalore, India, December 6–9, 2015, vol 9462. Springer, Berlin, pp 270–290

Guilley S, Hoogvorst P, Pacalet R (2004) Differential power analysis model and some results. In: Smart card research and advanced applications VI, IFIP 18th world computer congress, TC8/WG8.8 and TC11/WG11.2 Sixth international conference on smart card research and advanced applications (CARDIS), 22–27 August 2004, Toulouse, France, vol. 153. Springer, Berlin, pp 127–142

Heuser A, Picek S, Guilley S, Mentens N (2020) Lightweight ciphers and their side-channel resilience. IEEE Trans Comput 69(10):1434–1448. https://doi.org/10.1109/TC.2017.2757921

Heuser A, Picek S, Guilley S, Mentens N (2016) Side-channel analysis of lightweight ciphers: Does lightweight equal easy? In: Radio frequency identification and IoT security—12th international workshop, vol 10155. RFIDSec 2016, Hong Kong, China, November 30–December 2, 2016. Springer, Berlin, pp 91–104

Iwata T, Khairallah M, Minematsu K, Peyrin T (2019) Romulus v1. 2. In: Submission to NIST lightweight cryptography project

Kavut S, Baloğlu S (2016) Classification of 6 × 6 S-boxes obtained by concatenation of RSSBs. In: Lightweight cryptography for security and privacy—5th international workshop, vol 10098. LightSec 2016, Aksaray, Turkey, September 21–22, 2016. Springer, Berlin, pp 110–127

Kocher P, Jaffe J, Jun B (1999) Differential power analysis. In: Advances in cryptology—CRYPTO '99. 19th annual international cryptology conference, Santa Barbara, California, USA, August 15–19, 1999. Springer, Berlin, pp 388–397

Li H, Zhou Y, Ming J, Yang G, Jin C (2020) The notion of transparency order, revisited. Comput J 63(12):1915–1938. https://doi.org/10.1093/comjnl/bxaa069

Maghrebi H, Portigliatti T, Prouff E (2016) Breaking cryptographic implementations using deep learning techniques. In: Carlet C, Hasan MA, Saraswat V (eds) Security, privacy, and applied cryptography engineering—6th international conference, SPACE 2016, Hyderabad, India, December 14–18, 2016, vol 10076. Lecture Notes in Computer Science. Springer, Berlin, pp 3–26

Mangard S (2004) Hardware countermeasures against DPA ? In: A statistical analysis of their effectiveness. Topics in Cryptology—CT-RSA 2004, vol 2964. The cryptographers' track at the RSA conference 2004, San Francisco, CA, USA, February 23–27, 2004. Springer, Berlin, pp 222–235

Martínez-Díaz I, Freyre-Echevarria A (2020) S-boxes with theoretical resistance against power attacks under Hamming leakage models. https://www.

researchgate.net/publication/344233977_S-boxes_with_theoretical_resistance_against_power_attacks_under_Hamming_leakage_models. Accessed 7 June 2021
Naito Y, Matsui M, Sakai Y, Suzuki D, Sakiyama K, Sugawara T (2019) Saeaes. In: Submission to NIST lightweight cryptography project
NIST (2021) Lightweight cryptography standardization process. https://csrc.nist.gov/projects/lightweight-cryptography. Accessed 7 June 2021
O'Flynn C, Chen ZD (2014) Chipwhisperer: an open-source platform for hardware embedded security research. In: Constructive side-channel analysis and secure design—5th international workshop, COSADE 2014, Paris, France, April 13–15, 2014, vol. 8622. Springer, Berlin, pp 243–260
Patranabis S, Roy DB, Chakraborty A, Nagar N, Singh A, Mukhopadhyay D, Ghosh S (2019) Lightweight design-for-security strategies for combined countermeasures against side channel and fault analysis in IoT applications. J Hardw Syst Secur 3(2):103–131. https://doi.org/10.1007/s41635-018-0049-y
Picek S, Batina L, Jakobovic D (2014) Evolving DPA-resistant Boolean functions. In: Parallel problem solving from nature—PPSN XIII—13th international conference, vol 8672. Ljubljana, Slovenia, September 13–17, 2014. Springer, Berlin, pp 812–821
Picek S, Papagiannopoulos K, Ege B, Batina L, Jakobovic D (2014) Confused by confusion: systematic evaluation of DPA resistance of various S-Boxes. In: Progress in cryptology—INDOCRYPT 2014—15th international conference on cryptology in India, vol 8885. New Delhi, India, December 14–17, 2014. Springer, Berlin, pp 374–390
Picek S, Yang B, Mentens N (2016) A search strategy to optimize the affine variant properties of S-Boxes. In: Arithmetic of finite fields—6th international workshop, vol 10064. WAIFI 2016, Ghent, Belgium, July 13–15, 2016. Springer, Berlin, pp 208–223
Prouff E (2005) DPA attacks and S-Boxes. In: Fast software encryption: 12th international workshop, vol 3557. FSE 2005, Paris, France, February 21–23, 2005. Springer, Berlin, pp 424–441
Rivain M, Prouff E, Doget J (2009) Higher-order masking and shuffling for software implementations of block ciphers. In: Cryptographic hardware and embedded systems—CHES 2009, vol 5747. 11th international workshop, Lausanne, Switzerland, September 6–9, 2009. Springer, Berlin, pp 171–188
Runlian Z, Yaping S, Yongzhuang W, Yingxin L (2020) A new automatic search method for cryptographic S-Box. J Comput Res Dev 57(7):1415. https://doi.org/10.7544/issn1000-1239.2020.20190537
Smith LN (2017) Cyclical learning rates for training neural networks. In: 2017 IEEE winter conference on applications of computer vision, WACV 2017, Santa Rosa, CA, USA, March 24–31, 2017, pp. 464–472. IEEE Computer Society, Piscataway, NJ. https://doi.org/10.1109/WACV.2017.58
Standaert F-X, Peeters E, Quisquater J-J (2005) On the masking countermeasure and higher-order power analysis attacks. In: International symposium on information technology: coding and computing (ITCC 2005), vol 1. Las Vegas, Nevada, USA,4–6 April 2005. IEEE Computer Society, Piscataway, NJ, pp 562–567
Valiveti A, Vivek S (2020) Second-order masked lookup table compression scheme. IACR Trans Cryptogr Hardw Embed Syst 2020(4):129–153. https://doi.org/10.13154/tches.v2020.i4.129-153
Wouters L, Arribas V, Gierlichs B, Preneel B (2020) Revisiting a methodology for efficient CNN architectures in profiling attacks. IACR Trans Cryptogr Hardw Embed Syst 2020(3):147–168. https://doi.org/10.13154/tches.v2020.i3.147-168
Zhang W, Ding T, Yang B, Bao Z, Xiang Z, Ji F, Zhao X (2019) Knot: algorithm specifications and supporting document. In: Submission to NIST lightweight cryptography project
Zhu Y, Reddi VJ (2017) Optimizing general-purpose CPUs for energy-efficient mobile web computing. ACM Trans Comput Syst 35(1):1–1131. https://doi.org/10.1145/3041024

Publisher's Note