EM-X-DL: Efficient Cross-device Deep Learning Side-channel Attack With Noisy EM Signatures

JOSEF DANIAL and DEBAYAN DAS, Purdue University, USA
ANUPAM GOLDER, Georgia Institute of Technology, USA
SANTOSH GHOSH, Intel Corporation, USA
ARIJIT RAYCHOWDHURY, Georgia Institute of Technology, USA
SHREYAS SEN, Purdue University, USA

This work presents a **Cross-device Deep-Learning based Electromagnetic (EM-X-DL) side-channel analysis (SCA)** on AES-128, in the presence of a significantly lower **signal-to-noise ratio (SNR)** compared to previous works. Using a novel algorithm to intelligently select multiple training devices and proper choice of hyperparameters, the proposed 256-class **deep neural network (DNN)** can be trained efficiently utilizing pre-processing techniques like PCA, LDA, and FFT on measurements from the target encryption engine running on an 8-bit Atmel microcontroller. In this way, EM-X-DL achieves >90% single-trace attack accuracy. Finally, an efficient end-to-end SCA leakage detection and attack framework using EM-X-DL demonstrates high confidence of an attacker with <20 averaged EM traces.

CCS Concepts: • **Security and privacy** → **Embedded systems security**; **Side-channel analysis and countermeasures;**

Additional Key Words and Phrases: Electromagnetic side-channel attacks, cross-device attack, deep learning, profiling attacks, end-to-end SCA

ACM Reference format:
Josef Danial, Debayan Das, Anupam Golder, Santosh Ghosh, Arijit Raychowdhury, and Shreyas Sen. 2021. EM-X-DL: Efficient Cross-device Deep Learning Side-channel Attack With Noisy EM Signatures. *J. Emerg. Technol. Comput. Syst.* 18, 1, Article 4 (September 2021), 17 pages.
https://doi.org/10.1145/3465380

1 INTRODUCTION

With the ever-increasing prevalence of embedded devices and the growth of the **Internet of Things (IoT)**, the security of these devices has become a major concern. Some of the most serious threats to the security of these devices are side-channel analysis (SCA) attacks. By analyzing physical leakage information regarding the power [22], timing [23], or electromagnetic

This work was supported in part by the National Science Foundation (NSF) under Grants CNS 17-19235, CNS 19-35573, and in part by Intel Corporation.

Authors' addresses: J. Danial, D. Das, and S. Sen, Purdue University, West Lafayette, Indiana, USA; emails: {jdanial, das60, shreyas}@purdue.edu; A. Golder and A. Raychowdhury, Georgia Institute of Technology, Atlanta, Georgia, USA; emails: anupamgolder@gatech.edu, arijit.raychowdhury@ece.gatech.edu; S. Ghosh, Intel Corporation, Hillsboro, Oregon, USA; email: santosh.ghosh@intel.com.

1550-4832/2021/09-ART4 $15.00
https://doi.org/10.1145/3465380

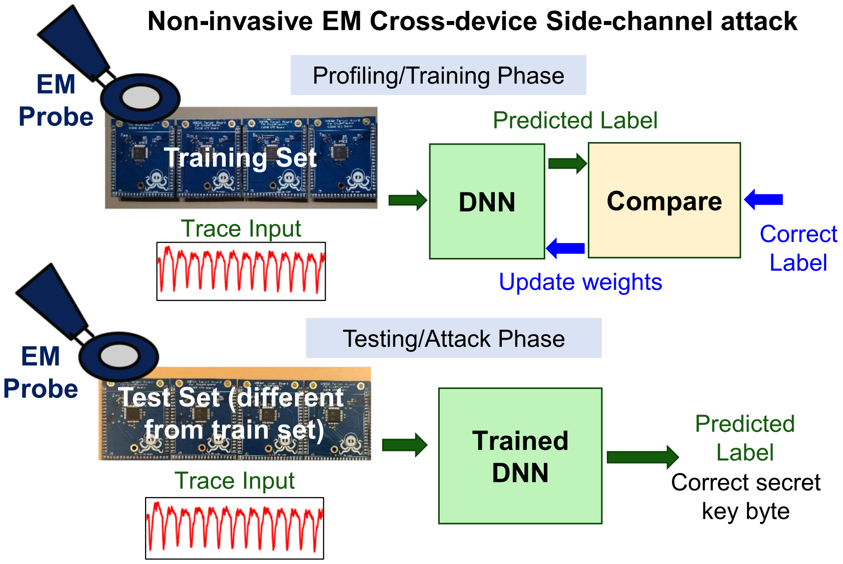

Fig. 1. Overview of cross-device EM profiled SCA. Profiled EM attacks are some of the most powerful side-channel attacks. By using EM measurements, the attack is non-invasive, an advantage over power attacks which require modifying the device under attack. Profiled attacks can reveal secret information using very few traces, but require a profiling phase. In a real attack, this profiling is done with separate devices completely under the attacker's control, then the attack is done on a new device. However, there exist variations from one instance of a device to another. These cross-device variations can cause problems if not accounted for.

(EM) signatures [2], cryptographic secrets can be extracted. Among the most powerful of these side-channel attacks are profiled attacks [8], and recently machine learning (ML) models have been shown to be very effective as both a defense, as in [35], and in the profiled SCA attack scenario using both power and EM measurements [18, 30].

Profiled attacks are incredibly powerful since far fewer victim measurements are needed to recover secret information compared to non-profiled attacks, reducing the time an attacker needs to have access to a device. Additionally, EM-based attacks are non-invasive, unlike power-based attacks, as the victim device does not need to be physically modified in any way to collect measurements. The basic setup of a profiled, EM-based side-channel attack is shown in Figure 1.

The main limitation of ML models for profiling SCA attacks is their portability to other target devices. Specifically, these models have been shown to work when the same device is used for both profiling and testing, however, in a real attack, the attacker would use a device to profile, then attack a separate, identical device. This issue of portability has recently been addressed for power ML SCA models on AES-128 in [4], [17], and [21], and also with a 3-class DNN attacking RSA implementations for EM SCA [7]. However, these works only consider high SNR scenarios, and with the introduction of SNR reducing countermeasures [10, 11, 14, 15], or low-cost, low-sensitivity EM probes [12, 16, 33], practical attacks must address the reality of low SNR trace measurements. In this work, we show a deep-learning-based cross-device SCA attack with low-SNR EM signatures.

1.1 Motivation

A 256-class DNN model that can be trained successfully (99% validation accuracy) [17] using raw time-domain AES-128 power traces for a particular microcontroller is rendered futile for low SNR EM SCA training even with traces collected from the same device (Figure 2(a)). Figure 2(b) shows

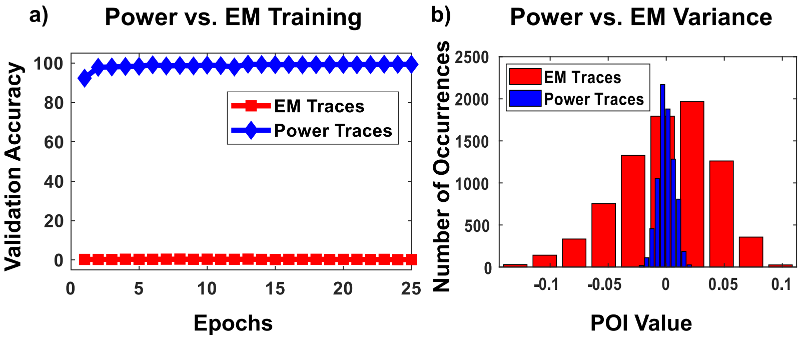

Fig. 2. (a) DNN training with high SNR raw power and low SNR raw EM traces. The model learns quickly from the power traces, but is unable to learn from the raw EM traces, demonstrating the need for new methods to train DNNs with EM traces. (b) The variance of a point of interest (time sample 103) for both power and EM traces, demonstrating significantly lower SNR of the EM traces.

the variance of a point of interest (POI, determined using the difference of means approach [8, 9]) across 10K EM and power traces. It clearly shows that the variation in the EM traces is much higher than the power traces, implying significantly lower SNR for the EM signatures. Indeed, when considering the side-channel SNR as defined in [26] as **SNR** $= \frac{VAR[Q]}{VAR[N]}$, with Q being the side-channel leakage and N being the noise, there is a large difference when comparing power and EM measurements. The side-channel SNR across a random selection of seven devices for power traces is 19.6 dB, while the SNR of equivalent EM traces is 3.1 dB, as is seen in Figure 3. Note that the SNR of a single device is comparable for both power and EM, but adding additional devices lowers the SNR drastically, due to the device-to-device variations. In fact, a majority of the lower SNR in the EM realm is due to *inter-device variations being more prominent in EM compared to power*, again looking at Figure 3. So, to solve the problem of portability, we need to take into account the inter-device variations [32]. To resolve all these issues, we utilize averaging to enhance the SNR, analyze different pre-processing techniques to reduce the dimensionality of the data, and develop an intelligent algorithm to choose the set of training devices for efficient profiling, given the need for more training devices to train a model due to the larger effect of inter-device variations. Finally, we also propose an end-to-end EM-X-DL attack framework to perform EM scanning and find the best point of leakage on an unseen target device. A combination of these techniques allows us to achieve >90% cross-device test accuracy.

1.2 Contribution

The specific contributions of this work are:

- This work presents a cross-device deep-learning-based EM SCA (EM-X-DL) on an AES-128 encryption engine using a 256-class DNN in a low SNR scenario, with ten devices for training and tested on a different set of ten test devices (Section 3).
- Effect of different pre-processing techniques including **principal component analysis (PCA)**, **linear discriminant analysis (LDA)**, **fast Fourier transform (FFT)**, spectrogram, on handling the portability issue is analyzed and compared, showing that the LDA is the most efficient approach to achieve maximum average cross-device key prediction accuracy of ~91.5% with minimum training time (Section 3.3).
- An algorithm for the optimal selection of the training devices is proposed, so that the number of training devices and thus the overall training time is minimized (Section 4, Algorithm 1).

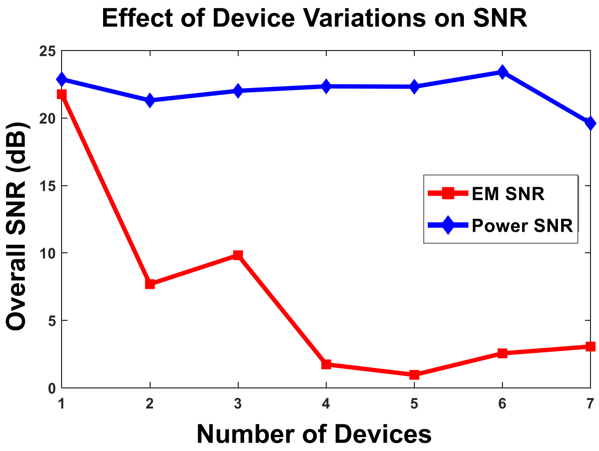

Fig. 3. Change in side-channel SNR as devices are added to the dataset for both power and EM. While the power SNR remains fairly high as additional devices are added, the EM SNR drops sharply, indicating a much larger effect of cross-device variations on side-channel leakage in the EM domain compared to power.

2 BACKGROUND & RELATED WORK

2.1 EM Side Channel Attacks

Since the inception of power SCA [22], a wide variety of attacks have been demonstrated, which can be broadly classified into non-profiled attacks like differential/correlational power/EM analysis (DPA, CPA, DEMA, CEMA) [5, 22], and profiled attacks, such as the statistical template attacks [8] and ML SCA attacks. While non-profiled attacks perform an attack in a single phase on a target device, profiled attacks consist of two phases, a profiling phase, to learn a leakage pattern and an attack phase, to attack with only a few traces, which practically operate on different devices. During the profiling stage, the attacker will collect traces from a "profiling" device identical to the victim device to build a model. During the attack, this model is then used to recover cryptographic secrets from the victim device.

2.2 ML-SCA Attacks

Template attacks have been shown [8] to be capable of recovering secret keys with a small number of traces, making them among the most powerful side channel attacks. More recently, supervised ML techniques have been used for profiling SCA [18]. Among these techniques, DNNs have been one of the most successful, defeating many common countermeasures, such as masking [20] and clock jitter [6]. Table 1 provides the summary of related works on profiling attacks. Till date, only one prior work [7] has focused on cross-device EM ML SCA attack using only one test device running RSA. Note that this attack required a 3-class DNN [7], whereas the proposed single-trace (averaged) EM-X-DL attack on AES-128 requires a 256-class DNN, and thus the effects of portability across devices is significantly more prominent. Additionally, AES measurements have significantly lower side-channel SNR compared to a public key algorithm such as RSA.

3 SIDE-CHANNEL ATTACK USING EM-X-DL

This section evaluates the single-trace (averaged) EM-X-DL attack on AES-128 using a 256-class DNN. For profiling the DNN, EM traces are collected from a set of ten training devices (8-bit Atmega microcontrollers) using the Chipwhisperer [29] platform, specifically the CW-Lite capture board, along with an off-the-shelf H-field sensor (10 mm loop diameter) and a 40 dB wideband

Table 1. Literature Review for Profiled-Attack Scenario

Profiled-Attack Scenario	Measurement Type	Profiling Method	Corresponding Article(s)
Same-Device	Power	TA	Chari et al. [2003] [8]
		SVM, RF	Bartkewitz et al. [2013] [18]; Lerman et al. [2014] [24]
		DNN	Maghrebi et al. [2016] [25]
	EM	TA	Chari et al. [2003] [8]
		DNN	Prouff et al. [2018] [30]
Cross-Device	Power	TA	Choudary and Kuhn [2018] [9]
		DNN	Das et al. [2019] [17]; Bhasin et al. [2019] [4]
	EM	TA	Montminy et al. [2013] [27]
		3-Class DNN	Carbone et al. [2019] [7] (RSA)
		256-Class DNN	**This Work*** **(AES-128)**

**First EM Cross-Device Deep-Learning Attack on a Symmetric Key Algorithm.*

amplifier. The efficient selection of the training devices is discussed in the subsequent section. For evaluating the attack, ten different devices are reserved separately and the cross-device (EM-X-DL) accuracy is reported as an average of these ten test devices.

3.1 Effect of EM Probe Choice

The EM probe used to collect both training and testing traces has an effect on the side-channel EM signals recorded. Two probes were considered: first a Langer probe with very high spatial sensitivity ($100 - \mu$ m diameter), and second a texbox probe with low spatial sensitivity ($10 -$ mm diameter). In order to estimate the leakage captured by each of the probes, the **test vector leakage assessment (TVLA)** [3] was used to measure side channel leakage, scanning over the surface of one of the 8-bit X-MEGA devices under attack. The results of these scans can be seen in Figure 7. As expected, the Langer probe finds high leakage in a very small area, while the larger probe detects leakage over a much larger area of the chip. Additionally, the larger probe detects a much higher level of leakage overall. Since the X-MEGA device is running a software implementation of AES-128, side-channel leakage is not highly localized, as it would be in a hardware implementation, and the high spatial sensitivity of the Langer probe does not provide a large benefit in rejecting algorithmic noise, as there is not a single register to target during an attack. For the rest of this work, results will be shown from the larger probe, as the leakage levels are already lower than power, and a variety of effects can be more easily investigated with the relatively higher leakage levels with the larger Tekbox probe—a t-value of 8 on the Langer probe vs. a t-value of 22 with the Tekbox probe, as seen in Figure 7.

3.2 DNN Architecture & Training

Figure 4 shows the architecture of the proposed 256-class **fully connected (FC)** DNN for the EM-X-DL attack. It should be noted that the EM traces captured using Chipwhisperer are time-synchronized and hence use of a convolutional layer is not necessary [21]. Three thousand time samples for each trace were collected from the 8-bit microcontrollers running AES-128, clocked at 7.37 MHz. Three thousand samples cover the entire AES encryption. Reducing the number of samples could affect accuracy if samples containing side-channel leakage are removed, but would

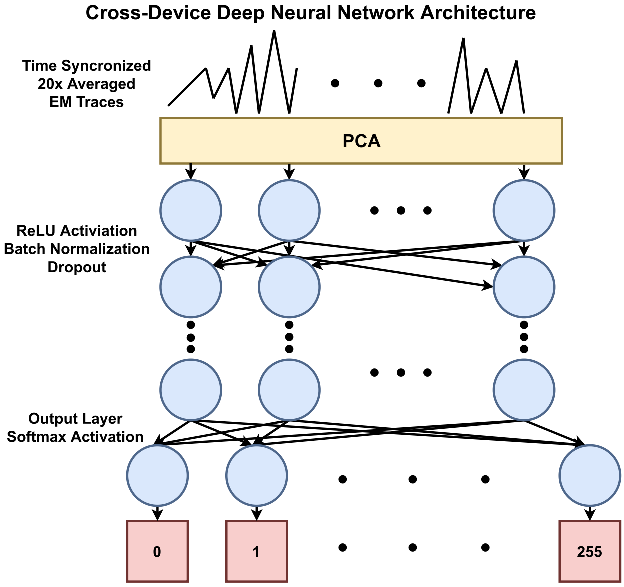

Fig. 4. Architecture of the proposed DNN. The network contains three dense layers, following each dense layer is a ReLU activation function, batch normalization, and finally a dropout layer. The final output layer provides the output class predictions - the key byte, and thus is size 256, and uses a softmax activation function.

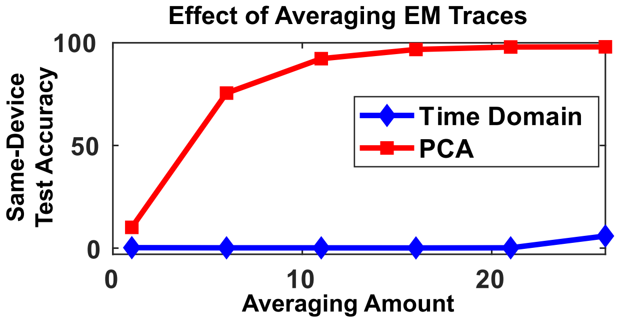

Fig. 5. Effect of averaging on the test accuracy of the 256-class DNN when using raw traces and PCA-transformed traces. Increasing averaging hardly allows the DNN to learn from the time-domain EM traces. With PCA used as a pre-processing step, averaging upto 20× smoothly increases the test accuracy to >99% for the same device.

be unlikely to otherwise—especially when PCA or LDA is used as a pre-processing method, as these methods will reduce the dimension of the trace, leaving only relevant side-channel information.

The DNN has a 3,000-neuron input layer, followed by three hidden layers with 100, 1,024, and 512 neurons, respectively, and finally the 256-neuron output layer. **Rectified Linear Unit (ReLU)** activation functions along with batch normalization and dropout used to achieve generalization are utilized for training the DNN. The Adam optimizer, with an initial learning rate of 0.005, which is halved whenever five consecutive training epochs pass without any validation accuracy improvement, is used for training. The effect of different hyperparameters is shown in Figure 6. A dropout of 0.45 is the most optimum for the first hidden layer (Figure 6(a)), while 1,024 hidden neurons for

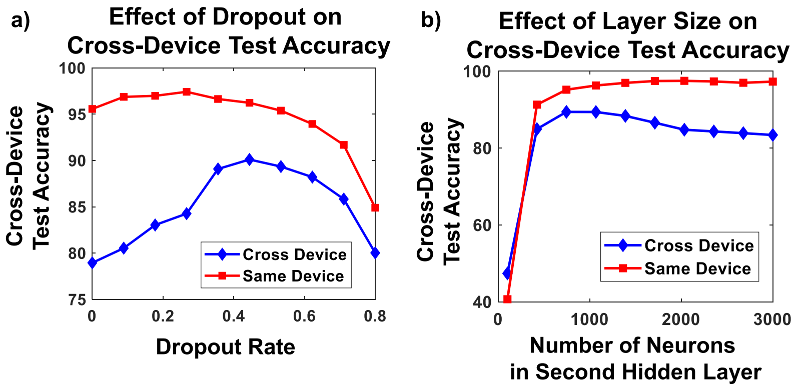

Fig. 6. Effect of hyperparameters on both same- and cross-device test accuracy for the PCA-DNN model. (a) Dropout between the first and second hidden layers helps prevent overfitting, maximizing cross-device test accuracy at a dropout rate of 0.45. (b) Layer size also demonstrates a similar trend, and reaches maximum cross-device test accuracy at ~1000 for the second hidden layer.

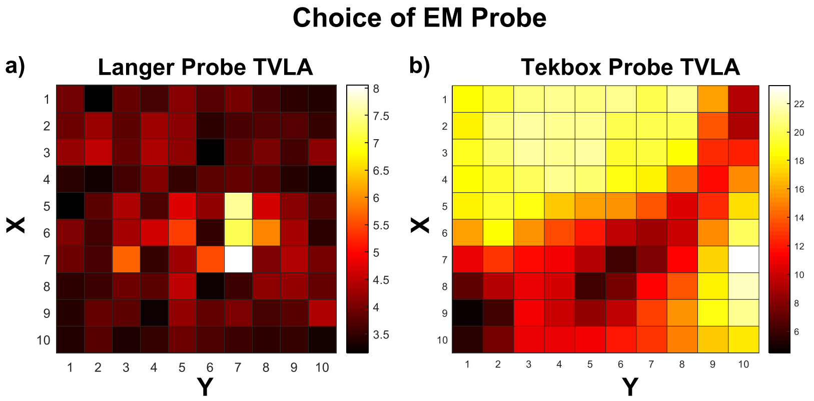

Fig. 7. Heatmaps created by performing a fixed vs. random TVLA on a 10 × 10 grid spanning the surface of the chip. (a) shows results obtained from a Langer ICR HH100-27 probe, while (b) shows results from a Tekbox probe. The leakage patterns of the two probes are quite different, which is not unusual as the Langer probe has a much higher spatial resolution. The Langer probe also measured the lower side-channel leakage overall, even at the maximum location.

the second hidden layer (Figure 6(b)) provides the maximum cross-device test accuracy without overfitting to the training devices. For all the results that follow, unless otherwise mentioned, the DNN is trained with ten devices for 100 epochs with a batch size of 64.

Now, as the raw EM traces collected from the ten training devices (100K traces each) are fed to the DNN classifier, the validation accuracy remains low (<1%) although training accuracy increases, even after 100 epochs. Figure 5 (blue curve) shows the effect of averaging on the same-device (test) accuracy. Even with 20× averaging, the time-domain traces shows a test accuracy of <1%, while a dimensionality reduction using PCA achieves >99% test accuracy for the same device. For cross-device attacks, the accuracy is lower, only 90% with 20× averaging and PCA. Figure 9 shows this result in terms of SNR, and shows the DNN's accuracy for lower levels of SNR as well (achieved by lowering the amount of averaging). As expected, the accuracy lowers with the SNR,

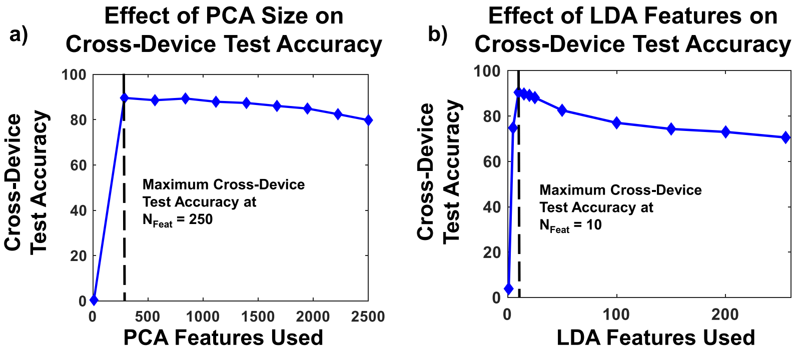

Fig. 8. PCA and LDA reach their respective peaks (250 and 10) with relatively few features compared to the size of the original traces (3000). As LDA features are chosen to maximize the class separation, while PCA maximizes variance, LDA is a more efficient technique for this higher dimensional data as it can train the DNN significantly faster.

following a similar pattern to Figure 5. Next, we will look into the effect of 20× averaging and different pre-processing strategies on the cross-device test accuracy. Note that, unless otherwise specified, cross-device test accuracy refers to the average key prediction accuracy of the EM-X-DL attack across all the ten test devices.

3.3 Single-trace Attack With Pre-processing

In the previous subsection, it was shown that the averaged time-domain EM traces (100K × 10 devices) do not train the DNN efficiently, while dimensionality reduction techniques like PCA have a significant impact in training the DNN. Here, we show results of using PCA [21] and LDA [32] on the time-domain EM traces, as well as the effects of frequency-domain-based processing (FFT, spectrogram [31, 36]) on the cross-device test accuracy for a DNN implemented using Tensorflow [1].

3.3.1 Dimensionality Reduction Using PCA & LDA. PCA transforms the input EM trace samples to their principal sub-space where individual features maximize the variance, while LDA achieves the same effect by maximizing the inter-class separation. As seen in Figures 8(a) and 8(b), the optimal number of features to use in these techniques is much lower than the dimensionality of the raw trace, around 250 in the case of PCA, and a mere 10 in the case of LDA. As shown in Figures 10(a) and 10(b), both of these techniques lead to roughly similar cross-device test accuracy, 91%. However, LDA is more efficient as it requires significantly lower training time ($<10\times$) than PCA to achieve the same level of accuracy. This can be seen in Figure 10, which shows that the accuracy of the LDA-based method stops increasing after ∼10 epochs, while in the PCA-based method, accuracy increases throughout all 100 epochs. In our case, this means that the time required to achieve iso-accuracy for LDA-based EM-X-DL is ∼100 seconds, while it is ∼1,000 seconds for PCA-based EM-X-DL. For both methods, as well as the others presented in this section, the time needed to recover the secret key is essentially the same and is low, as only 20 traces are needed, and prediction is done with a single evaluation of the neural network.

3.3.2 Frequency Domain Analysis Using FFT & Spectrogram. Using FFT on the time-domain averaged (20×) EM traces produces an EM-X-DL attack accuracy of ∼91% (Figure 10(b)), which is similar to PCA/LDA. However, it requires higher training time than both PCA and LDA, and hence is not the most efficient approach. Spectrogram combines both time- and frequency-domain

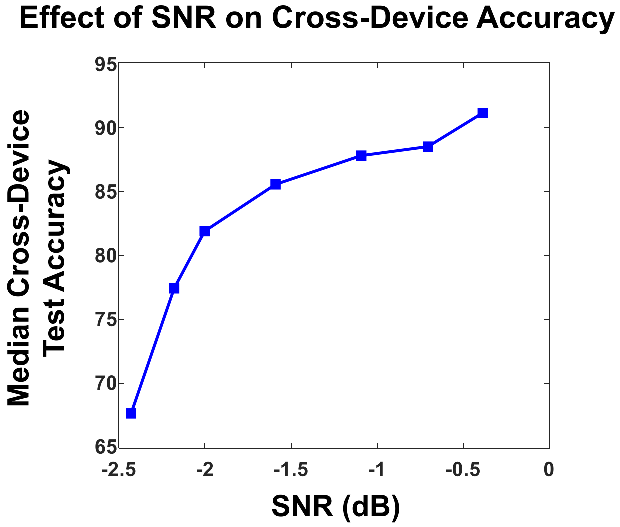

Fig. 9. Effect of side-channel SNR on the test accuracy of the 256-class DNN when using PCA-transformed traces. As expected, at higher SNR levels, the DNN achieves higher levels of cross-device test accuracy. Note that the limited SNR range is due to the use of avegraging to change SNR, while a majority of SNR reduction is a result of cross-device variations, as seen in Figure 3.

information and is naturally two-dimensional. Hence, a 2-D CNN [34] is used for the spectrogram, which achieves a cross-device test accuracy of 74.6% (Figure 10(b)).

4 EM-X-DL SCA: EFFICIENT SELECTION OF TRAINING DEVICES

As shown in the previous works [4, 17], the challenge of a ML SCA model being able to accurately classify traces collected from devices it has not been trained with, can be addressed by training with a variety of devices, so that the model does not overfit to the particular leakage pattern of one device. This remains true when using EM traces; however, many more devices are required to gain a high level of cross-device test accuracy, because, as seen in Figure 3, EM measurements are more sensitive to cross-device variations. Moreover, averaging clearly plays a key role, further increasing the number of traces required. Thus, it is of interest to be able to train using the smallest possible set of devices, reducing both the number of traces needed as well as the training time for the DNN. For this, two things must hold true: First, the choice of devices must affect the cross-device test accuracy for a given number of devices, and second, there must be a way of determining whether or not to include a device for training from a small sample of traces.

4.1 Cross-device Accuracy Variance

To address the first point, the effect of the subset, the EM-X-DL model is trained with a random subset of six devices, then tested against all the remaining fourteen devices. As shown in Figure 11, the average cross-device test accuracy can vary greatly even for a set of only six devices, with accuracy ranging from 10% to 75% for different six-device combinations. This shows that there are **subsets of training devices that can improve accuracy rather than simply adding more devices.** However, as there are a large number of possible subsets for a given size, an algorithm is necessary to choose one such subset which results in high cross-device test accuracy. Such an algorithm would then enable an attacker to gather quick measurements from a large set of devices,

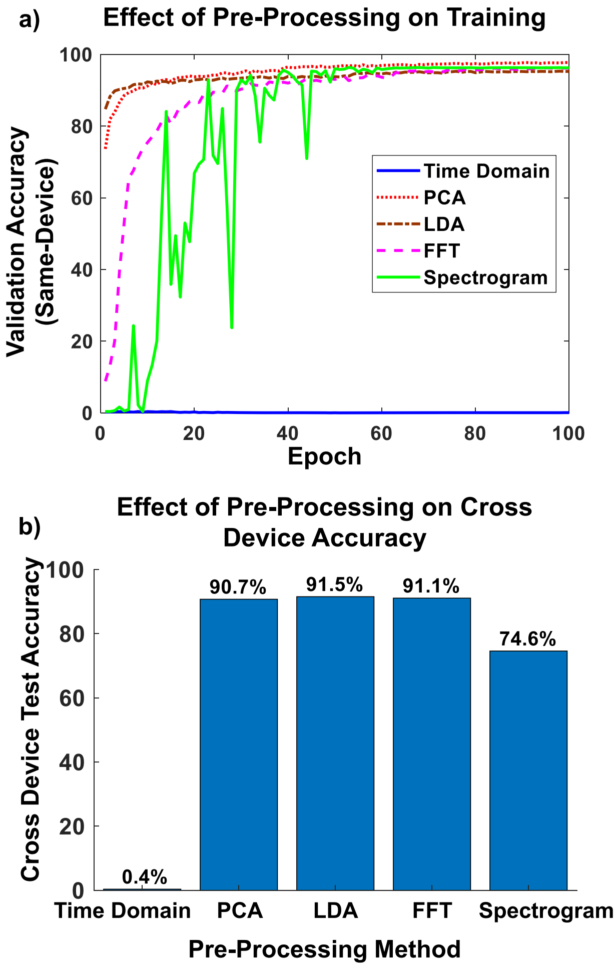

Fig. 10. Effect of the different pre-processing techniques on (a) the DNN training accuracy, (b) the cross-device attack (EM-X-DL) accuracy. While all the pre-processing techniques result in high validation (same-device) accuracy, PCA, LDA, FFT result in >90% cross-device test accuracy, while spectrogram yields 74.6% cross-device test accuracy.

and determine a small subset of devices to collect a large number of traces from, for training the DNN model.

4.2 Bivariate POI Based Device Selection

The proposed algorithm begins by identifying two points of interest (POIs) in the traces. This can be done through any POI identification technique, here POIs are chosen as time samples which have the highest difference of means (DOM). Once the top two POIs are found, the mean $\boldsymbol{\mu_i} = (\mu_{POI1}, \mu_{POI2})$ of this POI pair is calculated across all traces for each device. Then, to construct the subset of devices for training, one device is initially chosen arbitrarily, and additional devices are added as follows: The mean POI pair of all devices currently included in the training subset, $\boldsymbol{\mu_{train}}$ is calculated. Then, the next device is chosen such that $||\boldsymbol{\mu_i} - \boldsymbol{\mu_{train}}||_2$ is maximized, where i varies over all devices not already included in the training subset. In this way, at each step, the

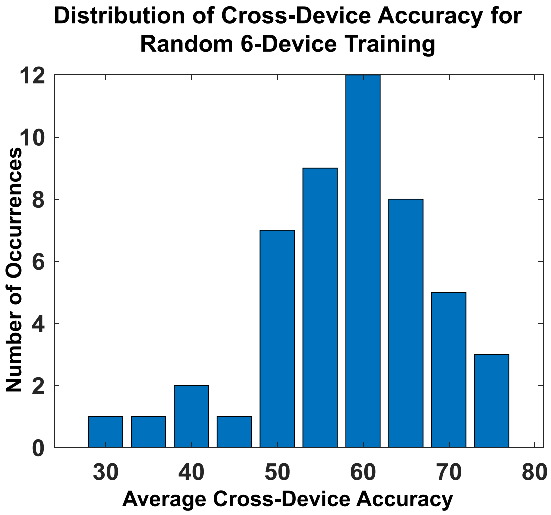

Fig. 11. Distribution of cross-device test accuracy of the 256-class PCA-DNN trained on random subsets of six devices. The mean accuracy is 60%; however, depending on the subset, it can vary significantly between 30–75%, highlighting the need for an intelligent selection of the training devices.

ALGORITHM 1: Algorithm for Device Selection

Input: Trace Samples from all Devices: TraceData, Number of Devices to select: nDev
Output: Subset of size nDev

```
for $dev$ = 1 : length(TraceData) do
    $\mu_1$ = mean(TraceData[dev][:,POI[1])
    $\mu_2$ = mean(TraceData[dev][:,POI[2])
    meanMap.append(dev, $(\mu_1, \mu_2)$)
end for
subset = [1]
for $i$ = 1 : nDev−1 do
    $\mu_{train}$ = mean(meanMap[subset])
    nextDev = argmax$_j$ ||$\mu_{train}$−meanMap[j][2]||
    subset.add(meanMap[nextDev][1])
    meanMap.remove(nextDev)
end for
return subset
```

device whose top two average POIs are furthest from the average POIs of the currently selected devices is added to the training set. This method is detailed in Algorithm 1, and depicted as a flow chart in Figure 12. The algorithm has a computational complexity of $O(mn + n^2)$ where m is the number of traces for each device, and n is the number of devices.

Figure 13 shows the 2-D bivariate normal distribution of the first three devices chosen using this algorithm, along with the total distribution of all devices. With these three devices, a large portion of the distribution spanned by all the devices is covered, revealing the successful operation of the algorithm. Importantly, this algorithm also provides the desired results during training, shown in Figure 14, as using this algorithm to choose the training devices gives higher cross-device (EM-X-DL) accuracy for any number of devices. Additionally, training with the devices closest to the

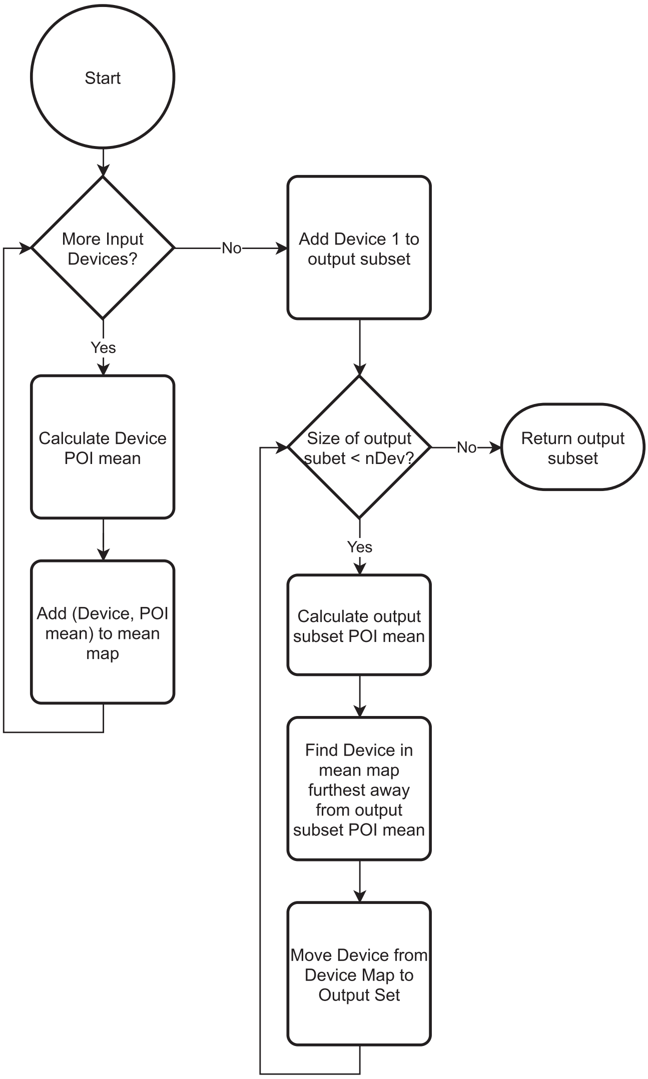
Start
More Input Devices?
No
Yes
Add Device 1 to output subset
Calculate Device POI mean
Add (Device, POI mean) to mean map
Size of output subet < nDev?
No
Return output subset
Yes
Calculate output subset POI mean
Find Device in mean map furthest away from output subset POI mean
Move Device from Device Map to Output Set

Fig. 12. Flow chart depiction of Algorithm 1. Each device is first characterized by a pair of mean POI values. Devices are then chosen for training by finding the device whose POI values are furthest from the mean POI of the currently selected devices.

current training set, as opposed to the furthest away, results in cross-device test accuracy significantly lower than the maximally different devices, and generally lower accuracy than randomly selected devices as well. These results were obtained with the proposed 256-class DNN, using 20× averaging and PCA-based pre-processing. From Figure 14, we also see that to attain a certain cross-device test accuracy, this algorithm requires between 20%–40% fewer training devices compared to random device selection.

5 EM LEAKAGE ASSESSMENT & ATTACK

Once the DNN model for the EM-X-DL SCA is trained, the main goal of an attacker is to break the secret key with minimum number of traces from an identical but unseen target device. This

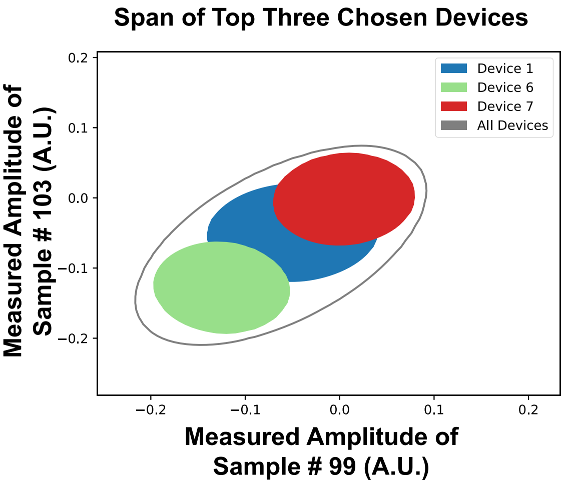

Fig. 13. Bivariate analysis of the first three devices chosen by Algorithm 1. The top three chosen devices already span a large portion of the distribution containing all devices.

section demonstrates an end-to-end attack strategy using the EM-X-DL model on a new device. By scanning the surface of the victim microcontroller and collecting traces at each point (seen in Figure 15(a)), the heatmap in Figure 15(b) was created by classifying the traces and determining the test accuracy for each point. As all training traces were collected from the same location (with maximum leakage on the chip evaluated using test vector leakage assessment (TVLA)), as expected the accuracy is highest in this region, then drops off sharply further from the measurement point. Figure 15(c) shows the minimum traces to disclosure (MTD) from a CEMA attack over the same chip. Comparing this to the accuracy heatmap shows that the ML model can correctly classify traces that are collected from a location which has an MTD less than 250.

Now, in this virtual grid, to converge to the best location for the EM-X-DL attack on the new device, the attacker can query the EM-X-DL model with multiple averaged traces collected from the test device and **observe if the frequency of the highest predicted key byte is distinguishable from the next.** Should leakage be present, the correct key byte would be predicted more often than others. If leakage is not present, predictions would be split between several key values. Thus, the ratio between the first and second most commonly predicted values provides a measure of the attacker's confidence in the prediction. This effect is shown in Figure 15(d), which shows the five most common predictions for both a location of high leakage (1,2) (left) and low leakage (2,9) (right). Note that, with this prior knowledge of the heatmap, the attacker can also divide the chip into four quadrants (for this particular chip) and get the correct key from the leftmost quadrant with a very high confidence. The correct key byte should be distinguishable because when the DNN is predicting correctly, that is, when there is leakage that the DNN was trained on, the softmax function ($\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$) makes the highest guess much higher than the next, due to the exponential nature of the softmax function. That is, if the DNN cannot predict when leakage is not present, then all values will be close, and even with softmax, there will not be a large separation between the first and second keybytes.

6 REMARKS & CONCLUSION

This work showed a Cross-device Deep Learning based EM (EM-X-DL) SCA attack on a symmetric key encryption engine (AES-128) in a low SNR setting. Utilizing a 256-class DNN, averaged EM

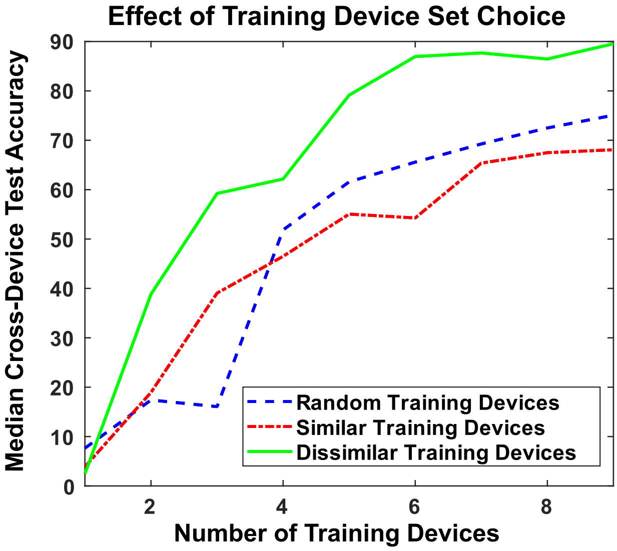

Fig. 14. Depending on the choice of devices used for training, cross-device test accuracy varies significantly. Choosing "dissimilar" devices by Algorithm 1 gives high accuracy, while choosing "similar" training devices yields a low cross-device test accuracy. Randomly selecting devices shows slightly higher test accuracies than choosing "similar" devices.

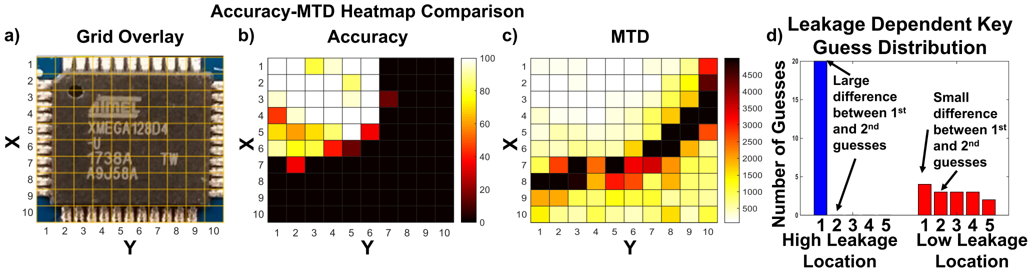

Fig. 15. (a) 10 × 10 virtual grid overlay of the chip. (b, c) Comparison of EM-X-DL model accuracy to CEMA-MTD. The ML model is able to predict with high accuracy in the region of the chip with low MTD values, however, when the MTD rises above 250, the model is unable to correctly predict the key values. (d) EM-X-DL model predictions on 20 samples from a high leakage location (1,1), and a low leakage location (9, 4) on a test device. At a location with high leakage, the frequency of the highest predicted key byte value is distinguishable from the next, demonstrating the high confidence of the attacker.

traces from 10 training devices along with dimensionality-reduction-based pre-processing (like LDA) the model achieves ~91.5% EM-X-DL single-trace (averaged) attack accuracy against another set of ten test devices. Table 2 summarizes the EM-X-DL attack accuracy for each of the different techniques studied in this article, while Table 3 compares EM-X-DL with other works. Note that the high accuracy of other works can be attributed to one of a number of factors. In the case of [7], the classifier is only 3-class, as the target algorithm is RSA, making training far simpler. In [17], the power side-channel is used, and has a considerably higher SNR compared to the EM side-channel considered in this work. Finally, Montminy et al. [27] uses an additional 500 traces from the test device to account for side-channel variations, easily an order of magnitude more than the number

Table 2. Cross-Device Attack Performance of Deep Learning-Based Methods for Different Pre-Processing Techniques

Preprocessing	*Cross – Device Accuracy* (%)		
Technique	Minimum	Average	Maximum
Time Domain	0.28	0.37	0.45
PCA	81.27	90.72	96.77
LDA	81.21	91.52	96.42
FFT	82.40	91.07	95.50
Spectrogram	30.53	74.58	94.02

Table 3. Comparison among EM-X-DL and Related Works

Corresponding Article	**Target Algorithm**	**Measurement Type**	**Number of Classes**	**Profiling Method**	**Traces Required**	**Accuracy**
Carbone et al. [2019] [7]	RSA	EM	**3**	DNN	15	99.91%
Das et al. [2019] [17]	AES-128	**Power**	256	DNN	1	99.9%
Montminy [2013] [27]	AES-128	EM	256	TA	**515**	99.6%
This Work	AES-128	EM	256	DNN	20	91.5%

of traces used in this work. An algorithm for efficient selection of training devices is proposed to speed up the profiling phase. Finally, an end-to-end attack using EM scanning is demonstrated showing that the attacker can detect the position of highest leakage on the chip using the proposed EM-X-DL model along with the secret key with high confidence.

For the future scope of this work, the end-to-end EM-X-DL attack can be more generalized by capturing traces from multiple locations across the chip, rather than a single location, for training the DNN. This would make the EM-X-DL attack much more efficient and faster as the attacker would be able to extract the key without having to detect one of the highest leakage locations on the chip.

Additionally, it was shown that the SNR of traces used for training and testing have a strong impact on the accuracy of the produced DNN as expected. This encourages development of countermeasures focused on reducing the SNR of side-channel signals, such as [13], [15], [19] and [28]. While such countermeasures can always fundamentally be defeated by collecting additional traces, by reducing the SNR significantly, collecting a sufficient number of traces from a large enough variety of devices becomes infeasible.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, Google Brain. 2016. TensorFlow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)*.

[2] Dakshi Agrawal, Bruce Archambeault, Josyula R. Rao, and Pankaj Rohatgi. 2003. The EM side-channel(s). In *CHES 2002*. 29–45. https://doi.org/10.1007/3-540-36400-5_4

[3] George Becker, J. Cooper, Elke DeMulder, Gilbert Goodwill, Joshua Jaffe, G. Kenworthy, T. Kouzminov, A. Leiserson, M. Marson, Pankaj Rohatgi, et al. 2013. Test vector leakage assessment (TVLA) methodology in practice. In *International Cryptographic Module Conference*, Vol. 1001. 13.

[4] Shivam Bhasin, Anupam Chattopadhyay, Annelie Heuser, Dirmanto Jap, Stjepan Picek, and Ritu Ranjan. 2019. Mind the portability: A warrior's guide through realistic profiled side-channel analysis. http://eprint.iacr.org/2019/661.

[5] Eric Brier, Christophe Clavier, and Francis Olivier. 2004. Correlation power analysis with a leakage model. In *CHES 2004*. 16–29.

[6] Eleonora Cagli, Cécile Dumas, and Emmanuel Prouff. 2017. Convolutional neural networks with data augmentation against jitter-based countermeasures. In *CHES 2017*.

[7] Mathieu Carbone, Vincent Conin, Marie-Angela Cornelie, François Dassance, Guillaume Dufresne, Cécile Dumas, Emmanuel Prouff, and Alexandre Venelli. 2019. Deep learning to evaluate secure RSA implementations. In *IACR Transactions on Cryptographic Hardware and Embedded Systems*. 132–161. https://doi.org/10.13154/tches.v2019.i2.132-161

[8] Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. 2003. Template attacks. In *CHES 2002*. 13–28. https://doi.org/10.1007/3-540-36400-5_3

[9] Marios O. Choudary and Markus G. Kuhn. 2018. Efficient, portable template attacks. *IEEE Transactions on Information Forensics and Security* 13, 2 (2018), 490–501. https://doi.org/10.1109/TIFS.2017.2757440

[10] Debayan Das, Josef Danial, Anupam Golder, Santosh Ghosh, Arijit Raychowdhury, and Shreyas Sen. 2020. Deep learning side-channel attack resilient AES-256 using current domain signature attenuation in 65nm CMOS. In *2020 IEEE Custom Integrated Circuits Conference (CICC)*. 1–4. https://doi.org/10.1109/CICC48029.2020.9075889 ISSN: 2152-3630.

[11] Debayan Das, Josef Danial, Anupam Golder, Nirmoy Modak, Shovan Maity, Baibhab Chatterjee, Donghyun Seo, Muya Chang, Avinash Varna, Harish Krishnamurthy, Sanu Mathew, Santosh Ghosh, Arijit Raychowdhury, and Shreyas Sen. 2020. 27.3 EM and power SCA-r AES-256 in 65nm CMOS through >350° current-domain signature attenuation. In *2020 IEEE International Solid- State Circuits Conference (ISSCC)*. 424–426. https://doi.org/10.1109/ISSCC19947.2020.9062997 ISSN: 2376-8606.

[12] D. Das, J. Danial, A. Golder, N. Modak, S. Maity, B. Chatterjee, D.-H. Seo, M. Chang, A. L. Varna, H. K. Krishnamurthy, S. Mathew, S. Ghosh, A. Raychowdhury, and S. Sen. 2020. EM and power SCA-r AES-256 through >350x current-domain signature attenuation and local lower metal routing. *IEEE Journal of Solid-State Circuits* (2020), 1–1. https://doi.org/10.1109/JSSC.2020.3032975

[13] Debayan Das, Santosh Ghosh, Arijit Raychowdhury, and Shreyas Sen. 2021. EM/power side-channel attack: White-box modeling signature attenuation countermeasures. *IEEE Design Test* (2021), 1–1. https://doi.org/10.1109/MDAT.2021.3065189

[14] D. Das, S. Maity, S. B. Nasir, S. Ghosh, A. Raychowdhury, and S. Sen. 2018. ASNI: Attenuated signature noise injection for low-overhead power side-channel attack immunity. *IEEE Transactions on Circuits and Systems I: Regular Papers* (2018), 1–12. https://doi.org/10.1109/TCSI.2018.2819499

[15] D. Das, M. Nath, B. Chatterjee, S. Ghosh, and S. Sen. 2019. STELLAR: A generic EM side-channel attack protection through Ground-Up root-cause analysis. In *2019 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. IEEE Computer Society, Los Alamitos, CA, 11–20. https://doi.org/10.1109/HST.2019.8740839

[16] D. Das, M. Nath, S. Ghosh, and S. Sen. 2020. Killing EM side-channel leakage at its source. In *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*. 1108–1111. https://doi.org/10.1109/MWSCAS48704.2020.9184657 ISSN: 1558-3899.

[17] Debayan Das, Anupam Golder, Josef Danial, Santosh Ghosh, Arijit Raychowdhury, and Shreyas Sen. 2019. X-DeepSCA: Cross-device deep learning side channel attack. In *DAC 2019*. ACM, 134:1–134:6. https://doi.org/10.1145/3316781.3317934

[18] Timo Bartkewitz and Kerstin Lemke-Rust. 2013. Efficient template attacks based on probabilistic multi-class support vector machines. In *Smart Card Research & Advanced Applications*.

[19] Archisman Ghosh, Debayan Das, Josef Danial, Vivek De, Santosh Ghosh, and Shreyas Sen. 2021. 36.2 An EM/power SCA-resilient AES-256 with synthesizable signature attenuation using digital-friendly current source and RO-bleed-based integrated local feedback and global switched-mode control. In *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, Vol. 64. 499–501. https://doi.org/10.1109/ISSCC42613.2021.9365978 ISSN: 2376-8606.

[20] R. Gilmore et al. 2015. Neural network based attack on a masked implementation of AES. In *HOST 2015*. 106–111. https://doi.org/10.1109/HST.2015.7140247

[21] Anupam Golder, Debayan Das, Josef Danial, Santosh Ghosh, Shreyas Sen, and Arijit Raychowdhury. 2019. Practical approaches toward deep-learning-based cross-device power side-channel attack. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 27, 12 (Dec. 2019), 2720–2733. https://doi.org/10.1109/TVLSI.2019.2926324

[22] Paul Kocher, Joshua Jaffe, and Benjamin Jun. 1999. Differential power analysis. In *CRYPTO'99*. 388–397.

[23] Paul C. Kocher. 1996. Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In *CRYPTO'96*. https://doi.org/10.1007/3-540-68697-5_9

[24] Liran Lerman, Gianluca Bontempi, and Olivier Markowitch. 2014. Power analysis attack: An approach based on machine learning. *Journal of Applied Cryptography* 3 (2014). https://doi.org/10.1504/IJACT.2014.062722

[25] Houssem Maghrebi, Thibault Portigliatti, and Emmanuel Prouff. 2016. Breaking Cryptographic Implementations Using Deep Learning Techniques. http://eprint.iacr.org/2016/921.

[26] Stefan Mangard. [n. d.]. Hardware countermeasures against DPA A statistical analysis of their effectiveness. In *Topics in Cryptology CT-RSA 2004* (2004) *(Lecture Notes in Computer Science)*, Tatsuaki Okamoto (Ed.). Springer Berlin, 222–235.

[27] David P. Montminy, Rusty O. Baldwin, Michael A. Temple, and Eric D. Laspe. 2013. Improving cross-device attacks using zero-mean unit-variance normalization. *Journal of Cryptographic Engineering*. https://doi.org/10.1007/s13389-012-0038-y

[28] Mayukh Nath, Debayan Das, and Shreyas Sen. 2021. A multipole approach towards on- chip metal routing for reduced EM side-channel leakage. *IEEE Microwave and Wireless Components Letters* (2021), 1–1. https://doi.org/10.1109/LMWC.2021.3062809

[29] Colin O'Flynn and Zhizhang Chen. 2014. ChipWhisperer: An open-source platform for hardware embedded security research. In *COSADE 2014*. 243–260.

[30] Ryad Benadjila, Emmanuel Prouff, Rémi Strullu, Eleonora Cagli, and Cécile Dumas. 2018. Study of deep learning techniques for side-channel analysis and introduction to ASCAD database. In *IACR Cryptology ePrint Archive*, vol. 2018. 53. http://eprint.iacr.org/2018/053.

[31] Christian Rechberger and Elisabeth Oswald. 2005. Practical template attacks. In *Information Security Applications*. https://doi.org/10.1007/978-3-540-31815-6_35

[32] Mathieu Renauld, François-Xavier Standaert, Nicolas Veyrat-Charvillon, Dina Kamel, and Denis Flandre. 2011. A formal study of power variability issues and side-channel attacks for nanoscale devices. In *EUROCRYPT 2011 (Lecture Notes in Computer Science)*. Springer, 109–128. https://doi.org/10.1007/978-3-642-20465-4_8

[33] Dong-Hyun Seo, Mayukh Nath, Debayan Das, Baibhab Chatterjee, Santosh Ghosh, and Shreyas Sen. 2021. PG-CAS: Patterned-ground co-planar capacitive asymmetry sensing for mm-range EM side-channel attack Probe Detection. In *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. 1–5. https://doi.org/10.1109/ISCAS51556.2021.9401580 ISSN: 2158-1525.

[34] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*. http://arxiv.org/abs/1409.1556.

[35] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu. 2018. IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? *IEEE Signal Processing Magazine* 35, 5 (2018), 41–49. https://doi.org/10.1109/MSP.2018.2825478

[36] Guang Yang, Huizhong Li, Jingdian Ming, and Yongbin Zhou. 2019. Convolutional Neural Network Based Side-Channel Attacks in Time-Frequency Representations. In *Smart Card Research and Advanced Applications*. 1–17. https://doi.org/10.1007/978-3-030-15462-2_1

Received October 2020; revised March 2021; accepted May 2021